

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26 2005, 13:45:55 ; Search time 44 Seconds
(without alignments)
900.878 Million cell updates/sec

Title: US-10-721-553-2

Perfect score: 2764

Sequence: 1 MAPIQTQQRDEGHRPNH.....QEDGSEAAASDSSEADSDSD 531

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCITUS-COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2764	100.0	531	2	US-08-933-750C-9
2	2764	100.0	531	3	US-09-234-613-9
3	2764	100.0	531	4	US-09-647-143-2
4	595	21.5	115	4	US-09-513-999C-7407
5	266	9.6	481	4	US-09-248-796A-18683
6	234.5	8.5	1742	4	US-09-386-962C-4
7	234.5	8.5	1742	4	US-09-386-959-4
8	233.5	8.4	930	4	US-09-200-650B-3
9	226	8.2	918	4	US-09-200-650B-1
10	225.5	8.2	933	3	US-08-293-728-2
11	225.5	8.2	933	3	US-09-421-868-2
12	225.5	8.2	936	4	US-08-956-171B-5249
13	225.5	8.2	936	4	US-08-781-986A-5249
14	223.5	8.1	1315	4	US-09-200-650B-5
15	220.5	8.0	1259	4	US-09-949-016-10366
16	218	7.9	1166	4	US-09-200-650B-7
17	212.5	7.7	287	4	US-09-710-279-468
18	212.5	7.7	1092	4	US-09-147-405B-15
19	203.5	7.4	414	4	US-09-248-796A-19046
20	203	7.3	257	3	US-09-461-697-188
21	203	7.3	272	3	US-09-461-697-186
22	199.5	7.2	599	4	US-09-538-092-864
23	198.5	7.2	218	3	US-09-461-697-190
24	196	7.1	781	4	US-09-949-016-9773
25	195	7.1	231	3	US-09-461-697-194
26	195	7.1	232	3	US-09-461-697-192
27	195	7.1	764	4	US-09-538-092-944

28 191 6.9 598 4 US-09-538-092-1083
29 189 6.8 764 4 US-09-370-838-67
30 189 6.8 764 4 US-09-854-133-67
31 186 6.7 3135 1 US-08-323-170B-2
32 186 6.7 3135 3 US-08-954-441-2
33 185.5 6.7 1162 2 US-08-728-323A-2
34 185.5 6.7 1162 3 US-09-298-568-2
35 185.5 6.7 1162 4 US-09-410-393-2
36 185.5 6.7 1162 4 US-09-894-273-2
37 185 6.7 402 4 US-09-248-796A-18910
38 184 6.7 674 3 US-08-893-852A-1
39 183 6.6 486 4 US-09-710-279-788
40 180 6.5 487 4 US-09-386-962C-14
41 180 6.5 487 4 US-09-386-959-65
42 178.5 6.5 40 4 US-09-647-143-16
43 178.5 6.5 1269 4 US-09-949-016-7349
44 178.5 6.5 1269 4 US-09-949-016-7350
45 178 6.4 1444 4 US-09-949-016-9652

ALIGNMENTS

RESULT 1
US-08-933-750C-9
; Sequence 9, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PITUNOR01

107c

```
; CLONE: 98974
US-08-933-750C-9

Query Match      100.0%; Score 2764; DB 2; Length 531;
Best Local Similarity 100.0%; Pred. No. 4.2e-227;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPIQTQOAREGHRPNHRTLPERSGVVCRVKYCNLSLPDPDPKFTTYPDQNRVQ 60
DB 1 MAPIQTQOAREGHRPNHRTLPERSGVVCRVKYCNLSLPDPDPKFTTYPDQNRVQ 60
QY 61 YKATSLKQKHQHDLLTEPDGLVTIDLINPDYTRIDPNVLLDPADEKLEEEIQAPTSSKR 120
DB 61 YKATSLKQKHQHDLLTEPDGLVTIDLINPDYTRIDPNVLLDPADEKLEEEIQAPTSSKR 120
QY 121 SQQAKVVPWVRKTEYISTEFNRYGINSNEKPEVKIGSVKQFTFEEIYKDRDSQITAE 180
DB 121 SQQAKVVPWVRKTEYISTEFNRYGINSNEKPEVKIGSVKQFTFEEIYKDRDSQITAE 180
QY 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVDFPKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
DB 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVDFPKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
QY 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRQDEEMDYAPDDVDYKIAREYNWVK 300
DB 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRQDEEMDYAPDDVDYKIAREYNWVK 300
QY 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360
DB 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360
QY 361 LEAQEARKAQLNHEPEEEEEEMETEKEAGGDEBEQKGSSEKSGSEDEHSGSESER 420
DB 361 LEAQEARKAQLNHEPEEEEEEMETEKEAGGDEBEQKGSSEKSGSEDEHSGSESER 420
QY 421 EGDREASDKSGGDESEDEARARDEKEIFGSDADSEDDADSDDEDRGQAQGGSDN 480
DB 421 EGDREASDKSGGDESEDEARARDEKEIFGSDADSEDDADSDDEDRGQAQGGSDN 480
QY 481 DSDSGSGGGQSRSHRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531
DB 481 DSDSGSGGGQSRSHRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531

RESULT 3
US-09-647-143-2
; Sequence 2, Application US/09647143
; Patent No. 6680196
; GENERAL INFORMATION:
; APPLICANT: Batra, Surinder K.
; APPLICANT: Hollingsworth, Michael A.
; APPLICANT: University of Nebraska Board of Regents

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PITUNOR01
; CLONE: 98974
US-09-234-613-9

Query Match      100.0%; Score 2764; DB 3; Length 531;
Best Local Similarity 100.0%; Pred. No. 4.2e-227;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPIQTQOAREGHRPNHRTLPERSGVVCRVKYCNLSLPDPDPKFTTYPDQNRVQ 60
DB 1 MAPIQTQOAREGHRPNHRTLPERSGVVCRVKYCNLSLPDPDPKFTTYPDQNRVQ 60
QY 61 YKATSLKQKHQHDLLTEPDGLVTIDLINPDYTRIDPNVLLDPADEKLEEEIQAPTSSKR 120
DB 61 YKATSLKQKHQHDLLTEPDGLVTIDLINPDYTRIDPNVLLDPADEKLEEEIQAPTSSKR 120
QY 121 SQQAKVVPWVRKTEYISTEFNRYGINSNEKPEVKIGSVKQFTFEEIYKDRDSQITAE 180
DB 121 SQQAKVVPWVRKTEYISTEFNRYGINSNEKPEVKIGSVKQFTFEEIYKDRDSQITAE 180
QY 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVDFPKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
DB 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVDFPKMWINPCAQVIFDSDPAPKDTSGAAALEM 240
QY 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRQDEEMDYAPDDVDYKIAREYNWVK 300
DB 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRQDEEMDYAPDDVDYKIAREYNWVK 300
QY 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360
DB 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRAKAGVQSGTNALLVVKHRDMNEKE 360
QY 361 LEAQEARKAQLNHEPEEEEEEMETEKEAGGDEBEQKGSSEKSGSEDEHSGSESER 420
DB 361 LEAQEARKAQLNHEPEEEEEEMETEKEAGGDEBEQKGSSEKSGSEDEHSGSESER 420
QY 421 EGDREASDKSGGDESEDEARARDEKEIFGSDADSEDDADSDDEDRGQAQGGSDN 480
DB 421 EGDREASDKSGGDESEDEARARDEKEIFGSDADSEDDADSDDEDRGQAQGGSDN 480
QY 481 DSDSGSGGGQSRSHRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531
DB 481 DSDSGSGGGQSRSHRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531

RESULT 2
US-09-234-613-9
; Sequence 9, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
```

; TITLE OF INVENTION: No. 6880196el Gene That is Amplified and
; FILE REFERENCE: Overexpressed in Cancer and Methods of Use Thereof
; CURRENT APPLICATION NUMBER: US/09/647,143
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US99/06633
; PRIOR FILING DATE: 1999-08-26
; PRIOR APPLICATION NUMBER: 60/079,649
; PRIOR FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FaastSeq for Windows Version 3.0
; SEQ ID NO: 2
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-647-143-2

Query Match 100.0%; Score 2764; DB 4; Length 531;
Best Local Similarity 100.0%; Pred. No. 4.2e-227; Indels 0; Gaps 0;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAPTIQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFPDQNRVQ 60
Db 1 MAPTIQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFPDQNRVQ 60
Qy 61 YKATSLKQKHDLTPDGLVGTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Db 61 YKATSLKQKHDLTPDGLVGTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Qy 121 SQQAAKVPVWRKTEYISTEFNRYGISNEKPEVKIGSVKQOFTFEEIYKDRDSQITAE 180
Db 121 SQQAAKVPVWRKTEYISTEFNRYGISNEKPEVKIGSVKQOFTFEEIYKDRDSQITAE 180
Qy 181 KTFEDAQKSIQHSYKPRVTPVNPVPPDFKMWINPCAVIFDSDPAPKDTSGAAALEM 240
Db 181 KTFEDAQKSIQHSYKPRVTPVNPVPPDFKMWINPCAVIFDSDPAPKDTSGAAALEM 240
Qy 241 MSQAMIRGMDDEGQFVAYFLPVEETLKKGRDQEEEMDYAPDDVYDKIAREYNNVVK 300
Db 241 MSQAMIRGMDDEGQFVAYFLPVEETLKKGRDQEEEMDYAPDDVYDKIAREYNNVVK 300
Qy 301 NKASGYENYFFIFREGDGVVYNELETRVRLSKERAKAGVQSGTNALLVVKHRDMNEKE 360
Db 301 NKASGYENYFFIFREGDGVVYNELETRVRLSKERAKAGVQSGTNALLVVKHRDMNEKE 360
Qy 361 LEAQEARKAQLENHEPEEEEEEMETEKEAGGSDEEKEGSSSEKESSEHSGSSESR 420
Db 361 LEAQEARKAQLENHEPEEEEEEMETEKEAGGSDEEKEGSSSEKESSEHSGSSESR 420
Qy 421 EGGDRDEASDKSGSEDESSDEARAARDKBEIFGSDADSDDEDDDRGQAGGSDN 480
Db 421 EGGDRDEASDKSGSEDESSDEARAARDKBEIFGSDADSDDEDDDRGQAGGSDN 480
Qy 481 DSDSGSGGGRSRRHSASAPFPFSGSHSAQEGSEAAASDSEADSDSD 531
Db 481 DSDSGSGGGRSRRHSASAPFPFSGSHSAQEGSEAAASDSEADSDSD 531

RESULT 4
US-09-513-999C-7407
; Sequence 7407, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 7407
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 25
; OTHER INFORMATION: Xaa=Glu or Gly
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 26
; OTHER INFORMATION: Xaa=Arg or Ser
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 110
; OTHER INFORMATION: Xaa=Glu or Gly
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 114
; OTHER INFORMATION: Xaa=Ala or Gly
US-09-513-999C-7407

Query Match 21.5%; Score 595; DB 4; Length 115;
Best Local Similarity 96.5%; Pred. No. 3.2e-43;
Matches 111; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 MAPTIQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFPDQNRVQ 60
Db 1 MAPTIQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFPDQNRVQ 60
Qy 61 YKATSLKQKHDLTPDGLVGTIDLINPDTYRIDPNVLLDPADEKLEEEIQAP 115
Db 61 YKATSLKQKHDLTPDGLVGTIDLINPDTYRIDPNVLLDPADEKLEEEIQAP 115

RESULT 5
US-09-248-796A-18683
; Sequence 18683, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18683
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18683

Query Match 9.6%; Score 266; DB 4; Length 481;
Best Local Similarity 22.0%; Pred. No. 2.8e-14;
Matches 110; Conservative 91; Mismatches 156; Indels 142; Gaps 20;
Qy 18 NSHRTL-PERSGVVCRVKYCNLSLPDIPDPKFIY----PFDQNRVQYKATSL-EKQHK 71
Db 16 SSKSLKPIQDYIAKYRTNNLPPLPKFIEYNTDPISTQOGEYLISSLFKXENF 75
Qy 72 HDLLTPD--LGVTIDLINPDTY-----RIDN---VLLDPADEKLEEE 110
Db 76 QNLMERIDQLGLDLNLRNGLSEDKNNESVGKLYNQLHPNDRALLRDAGIGKILKN 135
Qy 111 EIQAPTSKRSQOHAQVVPWRKTEYISTEFNRYGISNEKPEVKIG-----VSVKQ 161

Db 136 EPE-----VSFLRTEYIS-----DRPLSKGNNLNTATEEIKVKE 171
Qy 162 QFTTEEYIKDRDQSLTAIEKTFDAOKSI-----SOHYSKPRVTPVEVMPFDFQWLN 216
Db 172 KLSKDEHF-DADSOQNVESFTVAVESLYDLKNIKPKKKHLRAVNTWPLLPDTSMLDN 230
Qy 217 PCAQVIF-DSDPAPKDTSGAAALEMMSQAMIRGMDE-----EGNQFV 258
Db 231 VFINLRPMGASINRELNLKQOQOQOQNDKKFDEKJFDRALSLSLPKPIKLEGGWI 290
Qy 259 AYFL--PVEETLKKRKRQDOEEMDYAPDDVD--YKIAREYNWV-----KNKA 303
Db 291 SMYLLDATNTSTANDNDNEEQI-----NDLYELKHLTKKEQPINLLDEDEESLETYKFKY 346
Qy 304 SKYEENYFIFREGDV-----YNELTRVRLSKRA- 337
Db 347 TKNYDMTYQPFENEELAIKFVSDEIEDPVSKDNFKRKRMAYYYPIINGKIELKKHRAS 406
Qy 338 -----KAGVQSGTNALLVVKHRDMKELEAOBARKAOLENHEPEEEEEEMETEE 388
Db 407 TNSEINKPKERTYDGINFIL-----REPSTNELKRLDTINSEYDPMYEGDEEEEEEE 462
Qy 389 KEAGSDEEQEGSSSEKE 407
Db 463 EEPLEEEQQQSVETKEE 481

RESULT 6
US-09-386-962C-4
; Sequence 4, Application US/09386962C
; Patent No. 6635473
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLO
; FILE REFERENCE: P06335U2/BAS
; CURRENT APPLICATION NUMBER: US/09/386,962C
; CURRENT FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/098,443
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/117,119
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1742
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-386-962C-4

Query Match 8.5%; Score 234.5; DB 4; Length 1742;
Best Local Similarity 21.0%; Pred. No. 8.5e-11;
Matches 127; Conservative 88; Mismatches 240; Indels 149; Gaps 24;

Qy 48 FITYPPDQNRFFVQKATSLKQKHLDLLEPDLGVITDILNPDTYRIDP-----NVLLDP 102
Db 716 YVTLKDSNNRELQVTTDQSGHYQFDNLQNGT--YTFEFAIPDNYTPSPANNSTNDAIDS 773

Qy 103 ADEKLEEEIQAPTSSKRSQOQHAV-----VPMWRKTEYISTEFNRYGISNEKEPEVKIG 156
Db 774 DGERDGRKRVVAKGTINNADNMTVDTFGLYTPKYNVGVYVWEDTNKDGIDQDNEKGISG 833

Qy 157 VSV-----KQFT--EEETIY-----KDRDSQ 175
Db 834 VKVTLKKNKGDTIGTTTTDSNGKYEFTEGLENGDYTFIEFTPEGYTPTKQNSGSDGKDSN 893

Qy 176 ITALEKTFEDA-QKSIQSHYSKPRVTPVEVMPFDFQWLNPCAQVIFDSOP-----227
Db 894 GKTWTVVKADANKTIDSGFYKPTYN-----LGDY-VWEDTNKDGIDQDSEKGISGVK 945

Qy 228 -APKOTSGAA---ALEMMSQAMIRGMDEEGNQFVAYFLPVEETLKKRKKDOEEMDY- 281
Db 946 VTLKDKNGNAIGTTTTDASGHYQFKGL--ENGSTYVEFETPSGYTPTKANSQDITVDSN 1003

Qy 282 -----APDDVD---YKIAR-----EYNWNVYK--ASKGY-----307
Db 1004 GITTTGIINGADNLIDSGFYKTPKYSVGDYVWEDTNKDGIDQDNEKGISGVKVTLKDEK 1063
Qy 308 -----EENYFPIFREGD-GVYVNBLETRVRLSKERAKAGVQSGTNALLVVKHRDMN 357
Db 1064 GNIISITTTTIDNGKYQFDNLDSGNVYIIHFPEPEGTQTANS--NDD 1109

Qy 358 EKELEAQARKA-----OLENHEPEEEEEEMETERKEAGSGDEEQEGSSSEKEGSE 410
Db 1110 EKADGEGVRVTTIDHDDFSIDNGYFDDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 1169
Qy 411 DEHSGSEERE-EGDRDEASDK-SGSGDESEDEBARAKKEEIFGSDADEDDADSD 468
Db 1170 DSD 1229
Qy 469 EDRGAQOQSGSDSDSDSGNGGQGRSRSHRSASPPFSGSEHSAQBDGSEAAASDS-SEAD 527
Db 1230 DSD 1286

Qy 528 S0SD 531
Db 1287 S0SD 1290

RESULT 7
US-09-386-959-4
; Sequence 4, Application US/09386959
; Patent No. 6703025
; GENERAL INFORMATION:
; APPLICANT: PATTI, Joseph M.
; APPLICANT: FOSTER, Timothy J.
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: MULTICOMPONENT VACCINES
; FILE REFERENCE: P06333U1/BAS
; CURRENT APPLICATION NUMBER: US/09/386,959
; CURRENT FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: 60/098,439
; EARLIER FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1742
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-386-959-4

Query Match 8.5%; Score 234.5; DB 4; Length 1742;
Best Local Similarity 21.0%; Pred. No. 8.5e-11;
Matches 127; Conservative 88; Mismatches 240; Indels 149; Gaps 24;

Qy 48 FITYPPDQNRFFVQKATSLKQKHLDLLEPDLGVITDILNPDTYRIDP-----NVLLDP 102
Db 716 YVTLKDSNNRELQVTTDQSGHYQFDNLQNGT--YTFEFAIPDNYTPSPANNSTNDAIDS 773

Qy 103 ADEKLEEEIQAPTSSKRSQOQHAV-----VPMWRKTEYISTEFNRYGISNEKEPEVKIG 156
Db 774 DGERDGRKRVVAKGTINNADNMTVDTFGLYTPKYNVGVYVWEDTNKDGIDQDNEKGISG 833

Qy 157 VSV-----KQFT--EEETIY-----KDRDSQ 175
Db 834 VKVTLKKNKGDTIGTTTTDSNGKYEFTEGLENGDYTFIEFTPEGYTPTKQNSGSDGKDSN 893

Qy 176 ITALEKTFEDA-QKSIQSHYSKPRVTPVEVMPFDFQWLNPCAQVIFDSOP-----227
Db 894 GKTWTVVKADANKTIDSGFYKPTYN-----LGDY-VWEDTNKDGIDQDSEKGISGVK 945

Qy 228 -APKOTSGAA---ALEMMSQAMIRGMDEEGNQFVAYFLPVEETLKKRKKDOEEMDY- 281
Db 946 VTLKDKNGNAIGTTTTDASGHYQFKGL--ENGSTYVEFETPSGYTPTKANSQDITVDSN 1003

Qy 282 -----APDDVD---YKIAR-----EYNWNVYK--ASKGY-----307

Db 1004 GITTTGIINGADNTIHDGSGFYKPKYSGVDYVWEDTNKGDQDNEKGIQGVKVTLKDEK 1063
Qy 308 -----EENYFPIFREGD-GVYINELETRVRLSKERAKAGVQSGTNALLVVKHRDMN 357
Db 1064 GNIITTTTIDNGKQYQFNLDGNGVYIHFEKPEGWTQTANS-ND 1109
Qy 358 EKELEAQARKA-----QLENHEPEEBEEMETEKEAGGDERQEKSGSSEKGGSE 410
Db 1110 EKADAGEDVRVTITDHDGSDGSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 1169
Qy 411 DEHSGSEERE-EGDRDEASPK-SGSGDESEDEARAARKEIFGSDADSDSDSDSDSDSD 468
Db 1170 DSD 1229
Qy 469 EDRGOAQGSD 527
Db 1230 DSD 1286
Qy 528 SDSD 531
Db 1287 SDSD 1290

RESULT 8

US-09-200-650E-3
; Sequence 3, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Eidhinn, Deirdre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 3
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-200-650E-3

Query Match 8.4%; Score 233.5; DB 4; Length 930;
Best Local Similarity 20.5%; Pred. No. 4.2e-11;
Matches 130; Conservative 77; Mismatches 211; Indels 215; Gaps 25;
Qy 53 PQNRFFVQYKATSELEKQKHLLTPDLGVTIDLINPDTYRIDPNVLLDPADEKLLSEEI 112
Db 298 FEQVAFARKNATIDK-----TAYKMEVT---LGNDTY-----SEI 331
Qy 113 QAPTSSKESQQAQVWPMRKTEYISTE----- 140
Db 332 IDVYGNKKAQ-----PLISSTNYINNEEDLSRNTAYVQPKNTYTKQTFVNTLTGYKFN 385
Qy 141 -----FNRYGISNEK-----PEYKIGSVKQOFTBEIYKDRDSQITAEKTFDAQK 188
Db 386 PNAKNFKIYEDQNFVDSFTPDTSKLDVTDQF---DVIYSNDNKATATVD--LMKGQT 440
Qy 189 SISQHYKPRVTPVEMVFPDFKMWNPQAVIFDSDPAPKDTSGAAALEMMSQAMIRG 248
Db 441 SSNKQYIIQQV-----AYPDNSSTONGKIDYLDTDKTKYSWSN-----SYSNVG 486
Qy 249 MWDEEGNOFVAYEL---PVEETLKKRKEDQEE-----EMD----- 280
Db 487 SSTANGDO-KKYNLDGVVWEDTNKGDQANEKGIKGVYVILKDSNGKELDRITTTDENGK 545

Qy 281 -----YAP-----DDVVD-----YKI 291
Db 546 YQFTGLSNGTYSVEFSTPAGYTPPTANVGTDDAVIDSDGLTTGTGVIKADNMTLDSGFYKT 605
Qy 292 AR-----EYNWNVKNK-----ASKGY-----EENYFFIFREGDGV 321
Db 606 PKYSLGDYVWYDNGKDKRSTKGIKGVKVTQNEKGEVIGTTTETDENGKTRFDNLDG 665
Qy 322 YNNELETRVRLSKERAKAGV-QSGTNALLVVKHRDMNEKELEAQARKAQLN-HEPEEE 379
Db 666 KY-----KVIFEK---PAGLTQTGTNTTDDKADGGVDVTTTDDHDFLNGVYEEET 717
Qy 380 EREMETEEKAGGDEBEQKSGSSKEGSEDEHSGSESEEREGDRDEADKSGSGDEDS 439
Db 718 SD 777
Qy 440 SEDEARAARKEIFGSDADSD 499
Db 778 SD 836
Qy 500 ASPFPSPGSEHSAQEDGSEMAASDS-SEADSDSD 531
Db 837 DSD

RESULT 9

US-09-200-650E-1
; Sequence 1, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Eidhinn, Deirdre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1
; LENGTH: 918
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-200-650E-1

Query Match 8.2%; Score 226; DB 4; Length 918;
Best Local Similarity 22.6%; Pred. No. 1.8e-10;
Matches 125; Conservative 75; Mismatches 208; Indels 146; Gaps 23;
Qy 33 VKYCNSLPDIPP-DPKFITVFPQNFQVQYKATSELEKQKHLLTPDLGVTIDLIN--- 88
Db 276 VDYSNSNNTMPIADIK-----STNGDVAKAT-----YDILTCTYTFVTDYVNNKE 322
Qy 89 -----POTYRIDPNVLLDPADEKLLSE-----EIQAPTSSKRS 121
Db 323 NINGQFSLPFLTRAKAPKSGTYDANINI---ADEMFNFKITYNYVSSPIAGIDKPNGANIS 380
Qy 122 QQHAKVPMWRKTEYISTEFNRYGIGSNEKPEVIGSVKQOFTBEIYKDRDSQITAE- 180
Db 381 SLLIGVDVTSAGQNTYKQTVF-----VNPQRVLGNTVWYIKGYQDKI-ESSSGKVSATDT 434
Qy 181 --KTPE--DAQKISQHYKPRVTPV-EVMPVFPDFKMWNPQAVIFDSDPAPKDTSGA 235
Db 435 KLRIFEVNDTSKLSVYADPNDSNLKEVTDQKFNRIYEHHPNVASIKFGD----- 485
Qy 236 AALEMMSQAMIRGMDDEGNQFVAYFLPVEETLKKRKEDQEEEMDVPDDVTDYKIAREY 295

```
Db 486 --ITKYVVLVEGHYDNTG-----KNLKTQVIOENVDPVYNERDYSI---F 525
Qy 296 NMNVKNAKSGYEYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTNALLVVKHRD 355
Db 526 GWNNE-----VRYG-----GGADGSA----- 545
Qy 356 MNEKE-----LEAQARKAQL-----NHEPEEEEEEMETEKEAGSGSDEQEK 400
Db 546 VNPQPTPGPPVDPPEPPPEPTPPPEPPPEPPPEPPPEPPPEPPPEPPPEPPPEPP 605
Qy 401 GSGSEKGESEDEHSGSESERE-EGDRDEASDK-SGSGEDESSEDEARAARDKEEIFGSDA 458
Db 606 DSDSESDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 665
Qy 459 DSEDDADSDDERGQAQGGSDNDSCGNGGQRRSHRSASPPPSGSEHSAQEDGSEA 518
Db 666 DSDSESDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 724
Qy 519 AASDS-SEADSDSD 531
Db 725 SDSDSDSDSDSDSDSD 738

RESULT 10
US-08-293-728-2
; Sequence 2, Application US/08293728D
; Patent No. 6008341
; GENERAL INFORMATION:
; APPLICANT: Foster, Timothy J.
; APPLICANT: McDevitt, Damien L.
; TITLE OF INVENTION: The S. aureus Fibrinogen Binding Protein Gene
; FILE REFERENCE: 05344.105011
; CURRENT APPLICATION NUMBER: US/08/293,728D
; CURRENT FILING DATE: 1994-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 933
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-08-293-728-2

Query Match 8.2%; Score 225.5; DB 3; Length 933;
Best Local Similarity 21.4%; Pred. No. 2e-10;
Matches 103; Conservative 76; Mismatches 187; Indels 115; Gaps 18;

Qy 83 TIDLNP--DTYRIDPNVLDPADKLEEEIQAPTSSKRQOHAQVVPWPKRTEYISTE 140
Db 383 TIDQIDKNTYR--QTIYVNPNGDVI-----APVLT----- 413
Qy 141 FNRIGISNEKPEVKIGVSKQOFTEEIYK-----DRDSQITAEKTFEDAOKSISQHS 195
Db 414 -----GNLKPNTDSNALIDQNTSIKVYKVDNAADLSESYFVNPENFEDVTNSVNI 467
Qy 196 KPRVTPVEMVPVDPFKMWINPCAQVIFSDPAKDTSGAAALEMM-----SQAMIRGM- 249
Db 468 NPNQYKVEFNT--PDDQITTPYIVVNGHIDP---NSKGLALRSTLYGNSNIIWRMS 522
Qy 250 MDEEGNQFVAYFL-----PVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNWNVKNKA 303
Db 523 WDNE-----VAFNNGSGSGDIDKPVVPQDPPEGEIEPIED-----SDS 563
Qy 304 SKGYEENYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTNALL-----VVK 352
Db 564 DPGSDSG-----SDSNSDSDSGSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 601
Qy 353 HRDMNEKELEAQARKAQLNHEPEEEEEEMETEKEAGSGSEDEEKEGSSSEKESDE 412
Db 602 DSDSASDSDSASDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 661
Qy 413 HSGSESERE-EGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSDDER 471
Db 662 DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 720
Qy 304 SKGYEENYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTNALL-----VVK 352
Db 564 DPGSDSG-----SDSNSDSDSGSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 601
Qy 353 HRDMNEKELEAQARKAQLNHEPEEEEEEMETEKEAGSGSEDEEKEGSSSEKESDE 412
Db 602 DSDSASDSDSASDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 661
Qy 413 HSGSESERE-EGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSDDER 471
Db 662 DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 720
```

RESULT 12

US-08-956-171E-5249
; Sequence 5249, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 5249:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 936 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5249:
US-08-956-171E-5249
Query Match 8.2%; Score 225.5; DB 4; Length 936;
Best Local Similarity 21.4%; Pred. No. 2e-10;
Matches 103; Conservative 76; Mismatches 187; Indels 115; Gaps 18;
QY 83 TIDLNP--DTPRIDPNVLLDPADEKLEEEIOAPTSSKRSQOHAKVVPWVRKTEYISTE 140
DB 392 TIDQIDKNTNRYR--QTIYVNPSGDNVI-----APVLT-----422
QY 141 FNRYSISNEKEPVKIGVSVKQFTEEEIYK-----DRDSQITAEKTFEDAQKSIQSHYS 195
DB 423 -----GNLKPTDSNALIDQONTSIKYKVDNAADLSSEYFVNPENFEDVTNSVNTFF 476
196 KPRVTPVEVMPVDFDKWNPFCQVIFDSPAPKDTSGAAALEMM-----SQAMIRGM- 249
DB 477 NPQYKVFNT--EDDQITTYIVVNGHIDP-----NSKGDIALASTLYGYNLSIWRSMS 531
QY 250 MDEEGNQVAYFL-----PVEETLKRRKRODEEMDYAPDDVYDYKIAREYNNVKNKA 303
DB 532 WDNE---VAFNNGSGSDGIDKPVVPEQDPGEIEPIED-----SDS 572
QY 304 SKGVEENTFFIPREGDGYVYNELETRVLSKRRKAGVQSGTNALL-----VVK 352
DB 573 DPGSDG-----SDSNSDSDSGSDSDSTSDSGSDSDSDSAS 610

QY 353 HRDMNEKELEAQBARKAQLENHPEPEEEETEKEACGSGDEEEOEKSGSSKEGSEDE 412
DB 611 DSDSASDSDSASDSDSASDSDSDNDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 670
QY 413 HSGSESEERE--EGDRDEASDKSGSGEGESSEDEAARAARDKEEIFGSDADSDSDADSDDEDR 471
DB 671 DSD 729
QY 472 GQAGGSDNDSDSGSGGQGRSRSHSRKSPFPFGSGSHSAQEDGSEAAASD\$-SEADSDS 530
DB 730 SDS 788
QY 531 D 531
DB 789 D 789
RESULT 13
US-08-781-986A-5249
; Sequence 5249, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 5249:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 936 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-781-986A-5249
Query Match 8.2%; Score 225.5; DB 4; Length 936;
Best Local Similarity 21.4%; Pred. No. 2e-10;
Matches 103; Conservative 76; Mismatches 187; Indels 115; Gaps 18;
QY 83 TIDLNP--DTPRIDPNVLLDPADEKLEEEIOAPTSSKRSQOHAKVVPWVRKTEYISTE 140
DB 392 TIDQIDKNTNRYR--QTIYVNPSGDNVI-----APVLT-----422
QY 141 FNRYSISNEKEPVKIGVSVKQFTEEEIYK-----DRDSQITAEKTFEDAQKSIQSHYS 195
DB 423 -----GNLKPTDSNALIDQONTSIKYKVDNAADLSSEYFVNPENFEDVTNSVNTFF 476

Qy 196 KPRVTPVEMVPPDPFKMWINPCAQVIFDSDPAPKOTSGAAALEMM-----SQMIRGM- 249
Db 477 NPNQYKVEFNT--PDDQITTPYIIVVANGHIDP---NSKGLALRSLTYGNSNIWRSMS 531
Qy 250 MDEGNQFVAYFL-----PVEETLKKRKRDOEBENDYAPDDVVDYKIAREYNWNVKKA 303
Db 532 WDNE---VAFNNGSGGDGDKPVPEQDEPEIEPIED-----SDS 572
Qy 304 SKGYEENYPIFRGDDGVYNELETRVRLSKRRKAGVQSGTNALL-----VVK 352
Db 573 DPGSDSG-----SDSNSDSGSDSGSDSTSDSGSDSASDSAS 610
Qy 353 HRDMNKELEAQEARKAQLNHEPEEBEEMETEKEAGSGDEBEQKSGSSEKSGSEDE 412
Db 611 DSOSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSASDSAS 670
Qy 413 HSQSEBRE-EGDRDEASDKSGSGDESSDEARAAARDEEIEFGSDADSDDDADSDDEDR 471
Db 671 DSD 729
Qy 472 GQAQGGSDNDSGSGGGQGRSHRSASFPFSGSEHSAQEGSEAAASDS--SEADSDS 530
Db 730 SD 788
Qy 531 D 531
Db 789 D 789

RESULT 14
US-09-200-650E-5
; Sequence 5, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Eddihm, Deidre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1315
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-200-650E-5

Query Match 8.1%; Score 223.5; DB 4; Length 1315;
Best Local Similarity 21.7%; Pred. No. 4.9e-10;
Matches 121; Conservative 85; Mismatches 225; Indels 127; Gaps 23;

Qy 27 SGVVCVKYCNLSLPIFPDPKFTYFPDQNRFOYKATSLKQKHDLLEPDLGVITDL 86
Db 772 TGVV-----NGADMTLDSGF--YKTPKYNLGNVWEDTNKQKQDSTEGISGVTVL 823
Qy 87 INPD-----TYRID-----PNVLLDPADEKLEEBEIQAP 115
Db 824 KNENGEVLQTTKIDCKGYQFTGLENGYKVEPTEPSGYPTQVGSCTDEG-IDSNGTST 882
Qy 116 TSSKRSQOAHKV-----VPMWRKTEYSTENRYGISENEKPEVKIGVSVKQQTETEEIYK 170
Db 883 TGVIKDKNDTIDEGFYKPTYNLGDYVWEDTNKNGVQDKDEKGISGVTV-----TLK 934
Qy 171 DRDSQITALEKTEDAKQSISQ-----HYSKPRVTPVEMVPPDPFKMWINPCAQVIFDSD 226

Db 935 DENDKVLKVTVDENGKYQFTDLNNGTYKVBETPSGYTPT-----SVTSGN 981
Qy 227 PAPKOTSGAAALEMMSQAMIRGMDDEG-----NOFVAYFLPVEETLKKRKRDOEBE 277
Db 982 DTEKDSNGLTTTGVIKDA--DNMTLDSGFYKTPKYSGLGDIYVWY-----DSNKGKQDSTE 1034
Qy 278 ENDYAPDDVVDYKIAREYNWNVKKN--ASKGYEENYFFIFREGDGVYVYNELETRVLSKR 335
Db 1035 K-----GFKDVKVTL---LNEKGEVIGTKTDENKCYCFDNLDSKY-----KVIFEK- 1079
Qy 336 RAKAGV-QSGTGNALLVVKHRDMNKELEAQEARKAQLNHEPEEBEEMETEKEAGGS 394
Db 1080 --PAGLTGTGTNTTDDKADGGEVDVTTTDDHDFLDNGYVEETSD-----S 1126
Qy 395 DREQKSGSSEKSGSEDEHSGSEREREGRDEADSKSGSGEDESSEDARAAARDEKEIF 454
Db 1127 DSD 1184
Qy 455 GSDADSEDDADSDDRGOAQGGSDNDSGSGGGQGRSHRSASFPFSGSEHSAQED 514
Db 1185 DSD 1237
Qy 515 GSEAAASDS--SEADSDSD 531
Db 1238 SD 1255

RESULT 15
US-09-949-016-10366
; Sequence 10366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10366
; LENGTH: 1259
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10366

Query Match 8.0%; Score 220.5; DB 4; Length 1259;
Best Local Similarity 23.1%; Pred. No. 8.3e-10;
Matches 93; Conservative 57; Mismatches 160; Indels 93; Gaps 16;

Qy 171 DRDSQITALEKTEDAKQSISQHSYKPRVTPVEMVPPDPFKMWINPCAQVIFDSDPAPK 230
Db 297 DHDSGI-----GQNSDSKEYTDPGEK-----DPHNEV--DGDKTSK 331
Qy 231 DTSGAAALEMMSQAMIRGMDDEGNOFVAYFLPVBETLK-----KRQDEE 278
Db 332 SEENSA-----GIPDNGSQ-----RIEDTQKLNHRESKRVENRTKSETHA 374
Qy 279 MDYAPDDVVDYKIAREYNWNVKASKGYE-----ENYFFIFREGDGVYVNE--LETRV 330
Db 375 VGSQDKGIEIKGPPSSGNRNITKEVGKNEGKDGQHGMLKGNVKTQGEVNVIEGPG 434
Qy 331 RLSKRRAKAGVQSGTNALLVVKHRDMNKELEAQEARKAQLNHEPEEBE-----E 382
Db 435 QKSEPCNKVG-HSNTGS-----DSNSDGYDSYDFDKSMQGDPPNSDESNGNDANS 486
Qy 383 ENMETEKEAG-----GSDEQEKSGSSEKSGSEDEHSGSEREREGR-----DEASDK 431

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 13:48:56 ; Search time 174 Seconds
(without alignments)
1242.129 Million cell updates/sec

Title: US-10-721-553-2

Perfect score: 2764

Sequence: 1 MAPTQTQACREDGHRPNH.....QEDGSEAAADSDSDSD 531

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1826554 seqs, 40705358 residues

Total number of hits satisfying chosen parameters: 1826554

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/FCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2764	100.0	531	9	US-09-840-787-9
2	2764	100.0	531	17	US-10-721-553-2
3	2658.5	96.2	553	18	US-10-450-763-50041
4	2464	89.1	473	10	US-09-986-480-410
5	2464	89.1	473	17	US-10-863-332-410
6	622	22.5	133	18	US-10-450-763-50040
7	452	16.4	86	15	US-10-424-599-213359
8	375.5	13.6	571	15	US-10-424-599-213359
9	337.5	12.2	644	16	US-10-437-963-116147
10	283	10.2	475	18	US-10-450-763-49771
11	238.5	8.6	286	16	US-10-425-115-202470

12	237.5	8.6	1633	15	US-10-282-122A-70437	Sequence 70437, A
13	234.5	8.5	1742	15	US-10-615-383-4	Sequence 4, Appli
14	234.5	8.5	1742	16	US-10-690-184-4	Sequence 4, Appli
15	234.5	8.5	1742	16	US-10-689-082-4	Sequence 4, Appli
16	234	8.5	953	15	US-10-282-122A-44457	Sequence 44457, A
17	233.5	8.4	930	16	US-10-744-672-3	Sequence 3, Appli
18	233.5	8.4	930	17	US-10-744-616-3	Sequence 3, Appli
19	233.5	8.4	947	17	US-10-470-048B-86	Sequence 86, Appli
20	233	8.4	932	9	US-09-815-242-5578	Sequence 5578, Ap
21	233	8.4	932	9	US-09-815-242-12438	Sequence 12438, A
22	230.5	8.3	1385	15	US-10-282-122A-44324	Sequence 44324, A
23	230	8.3	877	15	US-10-282-122A-70428	Sequence 70428, A
24	229	8.3	1920	15	US-10-282-122A-71413	Sequence 71413, A
25	228	8.2	428	16	US-10-451-467A-448	Sequence 448, App
26	226	8.2	913	17	US-10-470-048B-63	Sequence 63, Appli
27	226	8.2	918	16	US-10-744-672-1	Sequence 1, Appli
28	226	8.2	918	17	US-10-744-616-1	Sequence 1, Appli
29	225.5	8.2	933	17	US-10-470-048B-93	Sequence 93, Appli
30	225.5	8.2	936	8	US-08-781-986A-5249	Sequence 5249, Ap
31	225.5	8.2	936	15	US-10-329-624-5249	Sequence 5249, A
32	224.5	8.1	194	18	US-10-450-763-41624	Sequence 41624, A
33	224	8.1	265	13	US-10-073-256-78	Sequence 78, Appli
34	223.5	8.1	1315	16	US-10-744-672-5	Sequence 5, Appli
35	223.5	8.1	1315	17	US-10-744-616-5	Sequence 5, Appli
36	223.5	8.1	1315	17	US-10-470-048B-124	Sequence 124, App
37	221.5	8.0	1349	9	US-09-815-242-5898	Sequence 5898, Ap
38	221.5	8.0	1349	9	US-09-815-242-13137	Sequence 13137, A
39	220.5	8.0	1253	14	US-10-363-798-2	Sequence 2, Appli
40	220	8.0	717	17	US-10-470-048B-58	Sequence 58, Appli
41	219	7.9	1021	9	US-09-815-242-5471	Sequence 5471, Ap
42	219	7.9	1021	9	US-09-815-242-12544	Sequence 12544, A
43	218	7.9	406	16	US-10-451-467A-262	Sequence 262, App
44	218	7.9	1166	16	US-10-744-672-7	Sequence 7, Appli
45	218	7.9	1166	17	US-10-744-616-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-09-840-787-9
; Sequence 9, Application US/09840787
; Patent No. US20020058264A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; Hillman, Jennifer L.
; Bandman, Olga
; Shah, Purvi
; Au-Young, Janice
; Yue, Henry
; Guegler, Karl J.
; Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/840,787
; FILING DATE: 23-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/518,865
; FILING DATE: <Unknown>

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: P1TUNOR01
; CLONE: 98974
; SEQUENCE DESCRIPTION: SEQ ID NO: 9 :
US-09-840-787-9

Query Match 100.0%; Score 2764; DB 9; Length 531;
Best Local Similarity 100.0%; Pred. No. 9.8e-167;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPTIQQAQREDGHRPNHRTLPERSGVVCRVKYCNSLPDIPDPKFIYPPDQNRVQ 60
Db 1 MAPTIQQAQREDGHRPNHRTLPERSGVVCRVKYCNSLPDIPDPKFIYPPDQNRVQ 60
Qy 61 YKATSLKQKHDLTEPDLGVTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Db 61 YKATSLKQKHDLTEPDLGVTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Qy 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITAE 180
Db 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITAE 180
Qy 181 KTFEDAQKSIQHSYKPRVTPEVMPVPDFPKWINPCAQVIFDSDPAPKDTSGAAALEM 240
Db 181 KTFEDAQKSIQHSYKPRVTPEVMPVPDFPKWINPCAQVIFDSDPAPKDTSGAAALEM 240
Qy 241 MSQAMIRGMWDEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
Db 241 MSQAMIRGMWDEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
Qy 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRAKAGVQSGTNALLVVKHRDMNEKE 360
Db 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRAKAGVQSGTNALLVVKHRDMNEKE 360
Qy 361 LEAQARKAOLENHEPREEEEEMETEKEAGGSDDEOEKGSSEKESGSEHSESER 420
Db 361 LEAQARKAOLENHEPREEEEEMETEKEAGGSDDEOEKGSSEKESGSEHSESER 420
Qy 421 EBGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGOAQGSDN 480
Db 421 EBGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGOAQGSDN 480
Qy 481 DSDSGNGGQGRSHRSRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531
Db 481 DSDSGNGGQGRSHRSRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531

RESULT 2
US-10-721-553-2
; Sequence 2, Application US/10721553
; Publication No. US20050032079A1
; GENERAL INFORMATION:
; APPLICANT: Batra, Surinder K.
; APPLICANT: Hollingsworth, Michael A.
; TITLE OF INVENTION: University of Nebraska Board of Regents
; TITLE OF INVENTION: Novel Gene That is Amplified and
; FILE OF INVENTION: Overexpressed in Cancer and Methods of Use Thereof
; FILE REFERENCE: UNMC6312H
; CURRENT APPLICATION NUMBER: US/10/721,553

```

```

; CURRENT FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: US/09/647,143
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US99/06633
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/079,649
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-721-553-2

Query Match 100.0%; Score 2764; DB 17; Length 531;
Best Local Similarity 100.0%; Pred. No. 9.8e-167;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPTIQQAQREDGHRPNHRTLPERSGVVCRVKYCNSLPDIPDPKFIYPPDQNRVQ 60
Db 1 MAPTIQQAQREDGHRPNHRTLPERSGVVCRVKYCNSLPDIPDPKFIYPPDQNRVQ 60
Qy 61 YKATSLKQKHDLTEPDLGVTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Db 61 YKATSLKQKHDLTEPDLGVTIDLINPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Qy 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITAE 180
Db 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITAE 180
Qy 181 KTFEDAQKSIQHSYKPRVTPEVMPVPDFPKWINPCAQVIFDSDPAPKDTSGAAALEM 240
Db 181 KTFEDAQKSIQHSYKPRVTPEVMPVPDFPKWINPCAQVIFDSDPAPKDTSGAAALEM 240
Qy 241 MSQAMIRGMWDEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
Db 241 MSQAMIRGMWDEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
Qy 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRAKAGVQSGTNALLVVKHRDMNEKE 360
Db 301 NKASKGYEENYFFIFREGDGVYNELETRVRLSKRAKAGVQSGTNALLVVKHRDMNEKE 360
Qy 361 LEAQARKAOLENHEPREEEEEMETEKEAGGSDDEOEKGSSEKESGSEHSESER 420
Db 361 LEAQARKAOLENHEPREEEEEMETEKEAGGSDDEOEKGSSEKESGSEHSESER 420
Qy 421 EBGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGOAQGSDN 480
Db 421 EBGDRDEASDKSGSGEDESSEDEARAARDKEEIFGSDADSEDDADSDDEDRGOAQGSDN 480
Qy 481 DSDSGNGGQGRSHRSRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531
Db 481 DSDSGNGGQGRSHRSRSASPPFSGSEHSAQEDGSEAAASDSEADSDSD 531

RESULT 3
US-10-450-763-50041
; Sequence 50041, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736

```

```

; SOFTWARE: Custom
; SEQ ID NO 50041
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (377)..(428)
; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,
; accession number BL00412D, p-value=9.633e-09, raw score of 16.54
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(553)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-50041

Query Match      96.2%; Score 2658.5; DB 18; Length 553;
Best Local Similarity 95.0%; Pred. No. 4.9e-160;
Matches 515; Conservative 6; Mismatches 10; Indels 11; Gaps 1;

Qy 1 MAPIQTQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYPPQNRFVQ 60
Db 12 MAPIQTQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYPPQNRFVQ 71
Qy 61 YKATSLKQKHDKHLLTEPDGLVITIDLPDTPYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Db 72 YKATSLKQKHDKHLLTEPDGLVITIDLPDTPYRIDPNVLLDPADEKLEEEIQAPTSKR 131
Qy 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQKQFTTEEIIYKORDSQITAE 180
Db 132 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQKQFTTEEIIYKORDSQITAE 191
Qy 181 KTFEDAQKS-----ISQHSKPRVTPVEVMPVPPDFKWINPACQVIFDSDPAP 229
Db 192 KTFEDAQKSVIEGLGWGEARISQHSKPRVTPVEVMPVPPDFKWINPACQVIFDSDPAP 251
Qy 230 KDTSGAAALEMMSQAMIRGMDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDY 289
Db 252 KDTSGAAALEMMSQAMIRGMDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDY 311
Qy 290 KIAREYNVNKSKAGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALL 349
Db 312 KIAREYNVNKSKAGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALL 371
Qy 350 VVKHRDMNEKELEQAQAKAQLENHEPEEEEEEMETEEKAGGSDDEQKSGSSEKEGS 409
Db 372 VVKHRDMNEKELEQAQAKAQLENHEPEEEEEEMETEEKAGGSDDEQKSGSSEKEGS 431
Qy 410 EDEHSGSESEEREGDRDEASDKSGGDESEDEARAAARDKEEIPGSDADSEDDADSDE 469
Db 432 EDEHSGSESEEREGDRDEASDKSGGQDDSSDYKARAARDKEEIPGSDADSEDDADSDE 491
Qy 470 DRGOAQGGSDNDSDSGRNGGGQRTSRHSRSASPPFSGSEHSAQEGSAAASDSEADSD 529
Db 492 DRGOAQGGSDNDSDSGRNGGGQRTSRHSRSASPPFSGSEHSAQEGSAAASDSEADSD 551

Qy 530 SD 531
Db 552 SD 553

RESULT 4
US-09-986-480-410
; Sequence 410, Application US/09986480
; Publication No. US2003002199A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 143 Human Secreted Proteins
; FILE REFERENCE: PS500P1
; CURRENT APPLICATION NUMBER: US/09/986,480
; PRIOR FILING DATE: 2001-11-08
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-50041
```

```

; PRIOR APPLICATION NUMBER: US 60/134,068
; PRIOR FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 456
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 410
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (405)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-986-480-410

Query Match      89.1%; Score 2464; DB 10; Length 473;
Best Local Similarity 99.8%; Pred. No. 8.3e-148;
Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MAPIQTQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYPPQNRFVQ 60
Db 1 MAPIQTQAOQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYPPQNRFVQ 60
Qy 61 YKATSLKQKHDKHLLTEPDGLVITIDLPDTPYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Db 61 YKATSLKQKHDKHLLTEPDGLVITIDLPDTPYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Qy 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQKQFTTEEIIYKORDSQITAE 180
Db 121 SQHAKVVPWPKTEYISTEFNRYGISNEKPEVKIGSVKQKQFTTEEIIYKORDSQITAE 180
Qy 181 KTFEDAQKSISQHSKPRVTPVEVMPVPPDFKWINPACQVIFDSDPAPKDTSGAAALEM 240
Db 181 KTFEDAQKSISQHSKPRVTPVEVMPVPPDFKWINPACQVIFDSDPAPKDTSGAAALEM 240
Qy 241 MSQAMIRGMDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNVNK 300
Db 241 MSQAMIRGMDEEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNVNK 300
Qy 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360
Db 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKE 360
Qy 361 LEAQEAQAKAQLENHEPEEEEEEMETEEKAGGSDDEQKSGSSEKEGSEDEHSGSESER 420
Db 361 LEAQEAQAKAQLENHEPEEEEEEMETEEKAGGSDDEQKSGSSEKEGSEDEHSGSESER 420
Qy 421 EBGDRDEASDKSGGDESEDEARAAARDKEEIPGSDADSEDDADSDEDDRSQ 473
Db 421 EBGDRDEASDKSGGDESEDEARAAARDKEEIPGSDADSEDDADSDEDDRSQ 473

RESULT 5
US-10-863-332-410
; Sequence 410, Application US/10863332
; Publication No. US20050064458A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 143 Human Secreted Proteins
; FILE REFERENCE: PS500P1
; CURRENT APPLICATION NUMBER: US/10/863,332
; CURRENT FILING DATE: 2004-06-09
; PRIOR APPLICATION NUMBER: US/09/986,480
; PRIOR FILING DATE: 2001-11-08
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-50041
```

FEATURE:
; NAME/KEY: SITE
; LOCATION: (405)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-863-332-410

Query Match 89.1%; Score 2464; DB 17; Length 473;
Best Local Similarity 99.8%; Pred. No. 8.3e-148;
Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAPTIQTAQREDGHRPNSHRTLPERSGVCRVKYCNLSLPDIPDPKFIITYPPDQNRVQ 60
DB 1 MAPTIQTAQREDGHRPNSHRTLPERSGVCRVKYCNLSLPDIPDPKFIITYPPDQNRVQ 60

QY 61 YKATSLKQKHDLTTPDLGVTTDLINPDTYRIDPNVLLDPADEKLEEEIOAPTSSKR 120
DB 61 YKATSLKQKHDLTTPDLGVTTDLINPDTYRIDPNVLLDPADEKLEEEIOAPTSSKR 120

QY 121 SQQAKVVPWNRKTEYISTENRNYGISNEKPEVKIGVSKOOFTEERIYKDRDSQITAE 180
DB 121 SQQAKVVPWNRKTEYISTENRNYGISNEKPEVKIGVSKOOFTEERIYKDRDSQITAE 180

QY 181 KTFEDAQKSIQRYSKPRVTPEVMPVFPDFKWINPCAQVIFDSDPAPKDTSGAAALEM 240
DB 181 KTFEDAQKSIQRYSKPRVTPEVMPVFPDFKWINPCAQVIFDSDPAPKDTSGAAALEM 240

QY 241 MSQAMIRGMDEENQFVAVLPVEETLKRRQRQDEEMDYAPDDVDYDKIAREYNWVK 300
DB 241 MSQAMIRGMDEENQFVAVLPVEETLKRRQRQDEEMDYAPDDVDYDKIAREYNWVK 300

QY 301 NKASKGYEENYFFIFRGDGVYVNELETRVLSKRAKAGVQSGTNALLVVKHRDWNKE 360
DB 301 NKASKGYEENYFFIFRGDGVYVNELETRVLSKRAKAGVQSGTNALLVVKHRDWNKE 360

QY 361 LEAQEARKAQLNHEPEEEEEETEKEAGGSDERQEKSGSSEKEGSDHSGSESER 420
DB 361 LEAQEARKAQLNHEPEEEEEETEKEAGGSDERQEKSGSSEKEGSDHSGSESER 420

QY 421 EBGDRDASDKSGEDESSEDEARAAARDKEEIFGSDADSDDDSDDRGQ 473
DB 421 EBGDRDASDKSGEDESSEDEARAAARDKEEIFGSDADSDDDSDDRGQ 473

RESULT 6
US-10-450-763-50040
; Sequence 50040, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60/736
; SOFTWARE: Custom
; SEQ ID NO 50040
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-450-763-50040

Query Match 22.5%; Score 622; DB 18; Length 133;
Best Local Similarity 64.4%; Pred. No. 7e-32;
Matches 130; Conservative 0; Mismatches 2; Indels 70; Gaps 1;

QY 273 RDQEEEMDYAPDDVDYDKIAREYNWVKKASKGYEENYFFIFREGDGVYVNELETRVRL 332
DB 273 RDQEEEMDYAPDDVDYDKIAREYNWVKKASKGYEENYFFIFREGDGVYVNELETRVRL 332

DB 1 ROOEEEMDYAPDDVDYDKIAREYNWVKKASKGY----- 35
QY 333 SKRRKAGVQSGTNALLVVKHRDMEKELEBAQARKAQLNHEPEEEEEETEKEAG 392
DB 36 -----EEEEEETEKEAG 50

QY 393 GSDEOEKSGSSEKEGSEDEHSGSEREGDRDEASDKSGGEDESSEDEARAAARDKEE 452
DB 51 GSVEOEKSGSSEKEGSEDEHSGSEREGDRDEASDKSGGEDESSEDEARAAARDKEE 110

QY 453 IFGSDADSDDDSDDDDRGQA 474
DB 111 IFGSDADSDDDSDDDDRGQA 132

RESULT 7
US-10-424-599-223174
; Sequence 223174, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223174
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_43557C.1.pap
US-10-424-599-223174

Query Match 16.4%; Score 452; DB 15; Length 86;
Best Local Similarity 100.0%; Pred. No. 2.4e-21;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LEKQKHDLTTPDLGVTTDLINPDTYRIDPNVLLDPADEKLEEEIOAPTSSKRQQOHA 125
DB 1 LEKQKHDLTTPDLGVTTDLINPDTYRIDPNVLLDPADEKLEEEIOAPTSSKRQQOHA 60

QY 126 KVPVWNRKTEYISTENRNYGISNEKP 151
DB 61 KVPVWNRKTEYISTENRNYGISNEKP 86

RESULT 8
US-10-424-599-213359
; Sequence 213359, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 213359
; LENGTH: 571
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_34688C.1.pap
US-10-424-599-213359

Query Match 13.6%; Score 375.5; DB 15; Length 571;
Best Local Similarity 27.2%; Pred. No. 1.6e-15;
Matches 128; Conservative 71; Mismatches 170; Indels 101; Gaps 19;

Qy 14 GHRPNHSTLP-----ER-----SGVCRVKYCNLSLPDPKFTYFPDQNRFPVQY 61
Db 150 GSRMGERSSTPLGAEVENRLKPTTFLCKLKFNLDPDPSAOPKLMASKKDDQVAKY 209
Qy 62 KATSLKQHKHDLITPDLGVITDILNPDTYRIDPNVL--LDPADKLLBEE-----IQA 114
Db 210 TITSLKMKYKPKLFEVDPGLDLDLDSVYN-PEVVRPPLAPDKELLRDEAVTPIKK 268
Qy 115 PTSSKRSQOHAKVPMWKRTEYISTEFNRYGISNEKPEVKIGSVKQOFTBEEI-----168
Db 269 DGIKRKERTPKGVAMLVKTVQVISP-----LSME-----STKQSLTEKQAKELREM 314
Qy 169 -----YKDRDSQITAEKTPEDAKSISQHYSKPRVTPVEVMPFDPKMWNPCA 219
Db 315 KGGRILOLNSRERQIREIEASF-AAKSDPVHATNKDLYPVEVMPPLLPDPDRYDQFV 373
Qy 220 QVIFSDP-----APKDTGAAALE---MMSQAMIRGMMDEEGNQFVAYFLPVEETLK 269
Db 374 VAFDNATATSEMAKXKDKSVDFESKAVKSVATGSDPANPEKFLATVWPAGELS 433
Qy 270 KRKRDOEEMDYAPDDVYDYKIAREYNNVKNKASKGYEENYFFIFREGDGVYNELETR 329
Db 434 KDIYDENEVSYS-----WIREFHMDVRGDDADD-PATFLVAFDSEARYL-PLPTK 483
Qy 330 VRLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEBEMEETEK 389
Db 484 LVLRKKRAKEG-RSG-----DEVRQCFVPARVTVRRSSVAAIERK 523
Qy 390 EAGGSDERQEGSSSEK-----EGSEDEHSGS---ESEREGDRDEASD 430
Db 524 DSG---VYTTSSKNSKRGSGLEMDGLEDQHRGAPHQDNYQSSGAEDYMSD 571

RESULT 9
US-10-437-963-116147
; Sequence 116147, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204566
; SEQ ID NO 116147
; LENGTH: 644
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_19676C.1.pap
US-10-437-963-116147

Query Match 12.2%; Score 337.5; DB 16; Length 644;
Best Local Similarity 25.3%; Pred. No. 4.7e-13;
Matches 109; Conservative 82; Mismatches 156; Indels 83; Gaps 17;

Qy 17 PMSHR---TLPRSGLVCEVKYCNLSLPDIPDPKFTYFPDQNRFPVQYKATSLKQHKHD 73
Db 228 PNAERFENRLKPTTFLCKHKNLDPDPSOLKWLPLNKDKORYTKRITSLEKNYIPK 287

Qy 74 LLTEPDLGVITDILNPDTYRIDPNVL-----LD 101
Db 288 MIVPEDLIGLFDLDDMSVYKIELSQIYVFLAAILRIRFFLAYLYAOTIFSTPPVQPPMA 347
Qy 102 PADEKLL-BEIIQAPTSS-----KRSQOHAKVPMWKRTEYI---STBFNRYEI---SNE 149
Db 348 PEDEELLRDDDEVLPVKDGIRKERTDKGMSLVKTVQVISPSTDAAKMSITEKQAKE 407
Qy 150 KPEVKIGSVKQOFTBEEIYKDRDSQITAEKTPEDAKSISQHYSKPRVTPVEVMPVFP 209
Db 408 RRESREG---RNTFLEN--INDREKQIKAIEDSGFR-AAKSRPVHQTCKRGMEAEWVLP 461
Qy 210 DFKWNPNCQAQVLFDSPPAPKDTSGAAAL------MMSQAMIRGMMDEEGNQFV 258
Db 462 DFRYDDQFVWVNFDPDPT-ADSEQVKNLRSERDECESRAVMKSLVNGSDPAKQKFL 520
Qy 259 AYFLPVEETLKKGRDOEEMDYAPDDVYDYKIAREYNNVKNKASKGYEENYFFIFREG 318
Db 521 AYVPSPHLSKDLDDDETDIQYS-----WLREYHWEVRGD-DKDDPTTYLVTF-DD 570
Qy 319 DGVYNELETRVRLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLN---HE 375
Db 571 DGAKYLPPLPTKLVLQKKAKEG-RSGDE----IEHFPVPSRITENLKRQSSVDDDL 625
Qy 376 PEREEEME 385
Db 626 PKHSRVEDMD 635

RESULT 10
US-10-450-763-49771
; Sequence 49771, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 49771
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (391)..(429)
; OTHER INFORMATION: ELEMENT TRANSPOSABLE INSERTION PROTEIN TRANSPOSITION DNA
; OTHER INFORMATION: domain identified by eMATRIX, accession number PD02455A, p-value:
; OTHER INFORMATION: 1.450e-25, raw score of 25.65
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (68)..(113)
; OTHER INFORMATION: Immunoglobulin domain identified by Pfam, accession name ig,
; OTHER INFORMATION: E-value=0.099, Pfam score of 10.6
US-10-450-763-49771

Query Match 10.2%; Score 283; DB 18; Length 475;
Best Local Similarity 87.7%; Pred. No. 9.2e-10;
Matches 57; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 328 TRVLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEBEMEETEK 387
Db 3 SRVLSKRRKAGVQSGTNALLVVKHRDMNEKELEAQEARKAQLNHEPEEEBEMEETEK 62
Qy 388 EXEAG 392

Qy 469 EDRGQAQGGSDNDSDSGNGGQGRSRSHRSASPPFPGSGSEHSAQEDGSEAAADS-SEAD 527
Db 1230 DSDAD 1286
Qy 528 SDSA 531
Db 1287 SDSA 1290

RESULT 15
US-10-689-082-4
; Sequence 4, Application US/10689082
; Publication No. US20040142348A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLOCOCCUS
; FILE REFERENCE: P06335US04/BAS
; CURRENT APPLICATION NUMBER: US/10/689,082
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: 09/386,962
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/098,443
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/117,119
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 1742
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-10-689-082-4

Query Match 8.5%; Score 234.5; DB 16; Length 1742;
Best Local Similarity 21.0%; Pred. No. 5.1e-06;
Matches 127; Conservative 88; Mismatches 240; Indels 149; Gaps 24;

Qy 48 PITVPPONRPFVQKATSLKQKHDLTPDGLVTVTLINPOTYRIDP-----NVLDP 102
Db 716 YVTLKDSNNRELQVTTDQSGHYQFNLQGT--YTVEFAIPDNYTPSPANNSTNDAIDS 773
Qy 103 ADEKLEEEIQARTSKRSQQAQV-----VPMRTEYISTEFNRYGIGSNEKPEVKIG 156
Db 774 DGERDGRKRVVAKGTINNADNMVTDTGFLTPKYNVGVYVWEDTNKGIQDDNEKIGISG 833
Qy 157 VSV-----KQOFT--BEEIY-----KDRDSQ 175
Db 834 VKVTLKKNKNGDTIGITTTTDSNGKVEFTGLENGDYTIEFETPEGYTPKQNSGDEGKDSN 893
Qy 176 ITALEKTFEDA-QKSIQSHYSKPRVTPVPMVPPDFPMINPCAVIFSDP-----227
Db 894 GTKTVTVKQADNKTIDSGFYKPTYN-----LGDY-VWEDTNKGIQDDSEKIGISGVK 945
Qy 228 -APKDTSGAA---ALEMMSQAMIRGMWDEGQFVAYFLPVEETLKKRRKQDEEMDY- 281
Db 946 VTLKKNKNGALGTTTDTASGHYQFGL--ENGSTVEFETPSGYTPKANSQGITVDSN 1003
Qy 282 -----APDDVYD---YKIAR---BYNMNVKXK-----ASKGY-----307
Db 1004 GIITTTGIINGADNLITDSGFYKTPKYSVGVYVWEDTNKGIQDDNEKIGISGVKVLKDEK 1063
Qy 308 -----BENYFFIFREGD-GVYNELETRVLSKRAKAGVQSGTNNALLVVKHRDMN 357
Db 1064 GNIISTTTTIDENGKYPDNLDSNGYIIHFKEKPEGMTQTANS-----NDD 1109
Qy 358 EKELEAQEAKA-----QLENHPEPEEEEEETEKEAGGSDDEQEKGSSSEKESGE 410
Db 1110 EKADGDEVRVTITDHDGDFIDNGYFDDSDSDSDSDSDSDSDSDSDSDSDSDSDSDADSDSDA 1169
Qy 411 DEHSGSESERE-EGDRDASDK-SGSGDESSDEAARAARDKEEIFGSDADSDSDSDSD 468
Db 1170 DS 1229

Qy 469 EDRGQAQGGSDNDSDSGNGGQGRSRSHRSASPPFPGSGSEHSAQEDGSEAAADS-SEAD 527
Db 1230 DSDAD 1286
Qy 528 SDSA 531
Db 1287 SDSA 1290

Search completed: September 26, 2005, 13:59:26
Job time : 176 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 13:44:30 ; Search time 42 Seconds
(without alignments)
1216.455 Million cell updates/sec

Title: US-10-721-553-2

Perfect score: 2764

Sequence: 1 MAPTIQQAQKEDGHRPNH.....QEDGSEAAASDSSEADSDSD 531

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79: *
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	645	23.3	425	2 T20261	hypothetical prote
2	335.5	12.1	547	2 C36828	unknown protein F1
3	280.5	10.1	457	2 T50233	probable DNA-direc
4	253	9.2	445	2 S44541	hypothetical prote
5	245.5	8.9	784	2 PN0009	neurofilament trip
6	241	8.7	792	2 T42963	hypothetical prote
7	236	8.5	590	2 A40437	glutamic acid-rich
8	234	8.5	953	2 C89824	hypothetical prote
9	230.5	8.3	1385	2 D89824	hypothetical prote
10	230	8.3	877	2 F90070	Clumping factor B
11	225.5	8.2	933	2 A51539	fibrinogen-binding
12	224	8.1	797	2 C36811	hypothetical prote
13	223.5	8.1	1315	2 T28679	fibrinogen-binding
14	221	8.0	1192	2 A1623	probable secreted
15	220	8.0	334	2 A54138	acidic repetitive
16	218	7.9	406	2 S38170	SRP40 protein - ye
17	218	7.9	1166	2 T28680	fibrinogen-binding
18	216	7.8	913	2 T52485	neurofilament prot
19	216	7.8	1110	2 I51116	NF-180 - sea lamp
20	214.5	7.8	1141	2 E89824	hypothetical prote
21	214	7.7	727	2 J51113	ribosomal transcri
22	214	7.7	2364	2 A56577	microtubule-associ
23	213.5	7.7	678	2 A54514	glutamic acid-rich
24	212.5	7.7	1092	2 T30214	fibrinogen-binding
25	211.5	7.7	665	2 B71609	hypothetical prote
26	211	7.6	985	2 D89852	fibrinogen-binding
27	208.5	7.5	798	2 I50479	neurofilament medi
28	208	7.5	765	2 S22314	transcription fact
29	203.5	7.4	606	2 S70358	centromere protein

ALIGNMENTS

RESULT 1

T20261

hypothetical protein C55A6.9 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T20261

R;Kershaw, J.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19243

A;Accession: T20261

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-425 <WIL>

A;Cross-references: UNIPROT:P90783; EMBL:Z81051; PIDN:CAB02869.1; GSPDB:GN00023; CESP:C:

A;Experimental source: clone C55A6

C;Genetics:

A;Gene: CESP:C55A6.9

A;Map position: 5

A;Introns: 14/2; 48/2; 90/3; 177/3; 381/1

Query Match 23.3%; Score 645; DB 2; Length 425;

Best Local Similarity 33.1%; Pred. No. 3.8e-26;

Matches 146; Conservative 96; Mismatches 165; Indels 34; Gaps 10;

Qy	24	PERSGVVCRVKYCNLSLPDIPDPKFIITYPF-DQNRVQYKATSLKQKHDLLETPDLGV	82
Db	14	PRKVDPMKPRFTNTVPDVFDAKFWTCFVPLGRFVFQPAIYRDYKHAVICDDMGL	73
Qy	83	TIDLINPDYTRIDP-NVLLDPADEKLEELIEIQAPTSKRSCQAHAKVVPWVKTEYISTEF	141
Db	74	NVLIDLKDYDEPIETEIDEKDNILLDDGAALKIAKRSQHSKLVFPWVKTEYISTEF	133
Qy	142	NRGISENEKEVIGSVKQFTTEELIKDRDQITAEIKTEFEDAQKSIQSHYKPRVTP	201
Db	134	NRFGVTADROETKLGYNLKNQVQVEDMYRDKQSIDAINKTFEDVRKPKVKEHYHKGKVA	193
Qy	202	VEVMPVFPDFKMINPCAQVIFSDPAKPDTSAGAAALMMMSQAMIRGMDEEGNQFVAYF	261
Db	194	VERSFVFPDFDHWKHLFAHVQFDGDTITTFEEDERQQARESSVIKAMEFDDQKFAAVF	253
Qy	262	LPVETLKKRQDEEMDYAPDDVDYKTAIREYNNWNKNAKSGYEENYFFIFREGDGV	321
Db	254	VPTIGCLTSFMDLLELRPFEDMKYEFLLSREYTFKMEHLPPR--DRDVFIMYHRNVVF	311
Qy	322	YNELETRVLSKRAKAGVQSGTNALLVVKHRDMNEKELEAQAKAQLNHE--PEEE	380
Db	312	QYNEVDCNVKMT-RKPKVALSRKSKULTTR----NFSELEQKDMNKREALYIQPKTRK	366
Qy	381	EEEME--TEEKEAGGSDEQEKSGSSEKESGSEDEHSGSESEREDGRDEADSKSGSEDE	438
Db	367	QEILEKIQEKKEGGD-----SSDQSSDDDDKPKQSR-----SDSSSDV	406

A:Molecule type: DNA
A:Residues: 1-445 <HOL>
A:Cross-references: UNIPROT:P38351; EMBL:X76053; NID:G600025; PIDN:CAA53642.1; PID:G4291
R:Brandt, T.; Christiansen, C.; Holmstroem, K.; Kalliesoe, T.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S46157
A:Accession: S46161
A:Molecule type: DNA
A:Residues: 1-445 <BRA>
A:Cross-references: EMBL:Z36148; NID:G536721; PIDN:CAA85243.1; PID:G536722; MIPS:YBR279W
R:Shi, X.; Finkelstein, A.; Wolf, A.J.; Wade, P.A.; Burton, Z.F.; Jaehning, J.A.
Mol. Cell. Biol. 16, 669-676, 1996
A:Title: Paf1p, an RNA polymerase II-associated factor in Saccharomyces cerevisiae, may
A:Reference number: JG6088; MUID:96140434; PMID:8552095
A:Accession: JG6088
A:Molecule type: DNA
A:Residues: 1-166,168-445 <SHI>
A:Experimental source: strain YJ3453
A:Accession: PC6031
A:Molecule type: DNA
A:Residues: 5-11;420-427 <SHZ>
C:Comment: This factor is a highly charged nuclear protein, and acts as a cofactor impor
C:Genetics:
A:Gene: SGD:PAF1
A:Cross-references: SGD:S0000483; MIPS:YBR279W
A:Map position: 2R
A:Note: this gene is located at the right arm of chromosome II
C:Superfamily: Saccharomyces cerevisiae hypothetical protein YBR279W
C:Keywords: nucleus
F:25-49/Region: PEST sequence
F:119-141/Region: nuclear location signal

Query Match 9.2%; Score 253; DB 2; Length 445;
Best Local Similarity 22.1%; Pred. No. 4.4e-06;
Matches 112; Conservative 96; Mismatches 184; Indels 114; Gaps 21;

Qy 23 LPERSGVCRVKNLSLDPIDPPDKFITYP-----FDQNRFOYKATSLKQKH 71
Db 1 MSKQEQYIAPIKYQNSLPVLPKPLVYPESPETNADSSQLNSLYIKTNVTLIQ-- 58
Qy 72 HDLLTEPLDGLTIDLI-----NPDTRIDPNVLLDPADEKLLEETQATSSKRS 121
Db 59 -----DEDLGHPVDLMPKPPGLNKLDSKLLYGFD-NVKLDKDDRIILLRD----PRIDRLT 108
Qy 122 QQAKVPMWPKTEIISTEENRYGISNEKPEKVGKSVKQFTEEEIYKDRDSQ----- 175
Db 109 KIDISKVTLFRTFVSNTIAAHNTSLKRRRL-----DQGSDDENLDV 154
Qy 176 ---ITAIEKTFEDAKSISQHSYKPRVTPVPMVPVFPDFKMWINPCAQVIFDSDPAKOT 232
Db 155 NHIISRVEGTFNKTDK--WQHPVKGVKQVKKWDLPLD-----TASMDQVYE----ILKF 203
Qy 233 SGAAALEMMSQMLIP-GM---MDEGNQFVAYFLPVEETLKKRKRDOEEEMDYAPDDVYD 288
Db 204 MGSASLDTKKKSLNTGIFRPEVEEEDWISMYATDHKDSAILNELEKMGDMDDDSHE 263
Qy 289 ---YKIAEYNNVKNKASKGYEENYFFIFREGDGV-YNLETKRVLSKRAKAGVQSG 344
Db 264 GKIKYKRIIDYMDKQVAKPWTE-LATRLNDKOGIAYKPLRSKIELRRRVNDIKP- 321
Qy 345 TNALLVVKH-----RDMNEKELEAQAKAQLN-----HEPEEEEEEMETEEL 389
Db 322 ---LVKEHDIQLNVLNLFNSTKEANIRDKLRFDPINFATVDEEDEDQEDVVK 377
Qy 390 BAGSDEEQKSGSSEKSGSDHSGSSEREGRDRDASDKSGSGEDESSEDARARD 449
Db 378 ESEG--DSKTEGSEGEENKDEETIKQEKENEQ-----DEENKQDENRAADT 422
Qy 450 KEETFGSDADSDDDSDDEDRGQAQ 475
Db 423 PET---SDAVHTEQKPEEEKETLOEE 445

RESULT 5
PN0009
neurofilament triplet M protein - Pacific electric ray (fragment)
C:Species: Torpedo californica (Pacific electric ray)
C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 09-Jul-2004
C:Accession: PN0009
R:Linial, M.; Scheller, R.H.
J. Neurochem. 54, 762-770, 1990
A:Title: A unique neurofilament from Torpedo electric lobe: sequence, expression, and lc
A:Reference number: PN0009; MUID:90155300; PMID:2106008
A:Accession: PN0009
A:Molecule type: mRNA
A:Residues: 1-784 <LIN>
A:Cross-references: UNIPROT:Q7LZ90
C:Comment: Neurofilaments are a subgroup of intermediate filaments which are expressed e
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; cytoskeleton; intermediate filament; nerve; phosphoprotein; tar
F:1-52/Region: serine-rich
F:53-84/Region: coil Ia
F:98-194/Region: coil Ib
F:217-367/Region: coil II
F:400-597/Region: glutamic acid-rich
F:598-674/Region: 6-residue repeats
F:675-784/Domain: carboxyl-terminal #status predicted <CTD>
F:616,622,628,634,640,646,652,658,670/Binding site: phosphate (Ser) (covalent) #status F

Query Match 8.9%; Score 245.5; DB 2; Length 784;
Best Local Similarity 20.5%; Pred. No. 2.1e-05;
Matches 122; Conservative 102; Mismatches 229; Indels 143; Gaps 23;

Qy 47 KPTTYPDQNRFOYKATSLKQ---HKHLLTEPLDGL-----VTIDLINPDYTR 93
Db 62 RFAGY-IDKVHYLEQQNKELEAETQAHKQKQVSHGLGVYDQREIRLSIEQVNOEKAQ 120
Qy 94 ID-PNVLLD-----PADEKLLEEE-----IQAPTSSKSGSOHAKV 128
Db 121 IQDLSVHLDDFORVGAFFDEALRDEPEATIRLVKKEETDSVYQADGGEKKAQSLQDEV 180
Qy 129 PMWRKTEYISTEENRYGISNEKPEKVGKSVKQFTEEEIYKDRDSQITAIHKTFEFAQK 188
Db 181 AFLR-----NNHEEV-ADLFAQIAQATQVTVKE-KDFLKTDTISALKEIRS 224
Qy 189 SISQHSYKPRVTPVPMVPVFPDFKMW-----INPCAQVIFDSDPAKOTSGAAALEMM 241
Db 225 QLECHSAKNMQQADE-----WPKCRYDKLNEAAEMNKDAIRAAREEIGSYRRQLQ 274
Qy 242 SQAM-----IRGM-----MDEGNQFVAYFLPVEETLKKRKRDOEEEM----DYA 282
Db 275 SKSIELESVSTKESLERQLTDIEDRHADVANYQETVQQLNELRGTKWENARHLREY- 333
Qy 283 PDDVYDYKIAEYNNVKNKASKGYEENYFFIFREGDGVYNN-----ELETRVLSKRRRA 337
Db 334 -QDLLNVKMALDIEIAAYRKLLGEESRYTF-SGTGPSIPYRSPRRLPAKVHKTKEVP 391
Qy 338 KAGVQSGTNALLVVKHRDMNEK-----ELEAQAKAQLNENHEPEEEEEEMETE----- 388
Db 392 KVKYQHKFVEEIIETETKVQDEKAEWGDIDLAEEVGGATWESPEDKEAEKVEEIVAT 451
Qy 389 -----KEAGSDEEQKSGSSEKSGSDHSGSSEREGRDRDASDKSGSGEDESSED 442
Db 452 VKGQVQAPPGAEAESESEAKKEEDEGVVEEEKKKE-EADDEEKEEKEDEGEADAEAGG 510
Qy 443 EARAARDKEEIF-----GSDADSEDDADSDDEDRGQAQGGSDNDSDSGNSGGQGR- 492
Db 511 ESRVVEKEVIVKVEQSKAHPGKDEVKEKEKEEKEEKEEKEEKEEKEEKEEKEEKEE 570
Qy 493 -----SRSHSRASAPFPFGSGSHSAQED-----GSEAAASDSEADSDSDS 530
Db 571 SKGKVEEKLTVTEKTEKATEDKVPREKPKQKEQKQIEEKKKEAKSKDEAKSKDEAK 626

RESULT 6
T42963

Qy	21	RTLPERSGVCRVKY---CNSLPOLPFPDPKITYFPDQNRVFOYKATSLKQKHDLITPEP	78
Db	7	RVLPQPPGTPQKTQBESEGTPEPEPKPTAPEE-----TELEVSUPPE-----EP	55
Qy	79	DLGVTIDLINPDYRIDPNVLLDPADEKLEEBEIQAPTS---SKRSQQHAKVVPMMRK---	133
Db	56	CVGKEVAAVTLGPGQTQETALTPT-----SLQAQVSVAPAEAHSSPRGVLTLWRKGV	108
Qy	134	-----TEYISTSTFNRYGISENPEVKIGVSVKQOFTTEEBEYKORDSQITAE-----	180
Db	109	EKVVPQPAHSRSPSONIAAGLESPPDQQAQIILGCGCTGG--SDEPSEPSAEADFGPGPW	166
Qy	181	--KTFE--DAOKSISQHYSKPRVT-----PVEVMVVF-----	208
Db	167	LLRWFENLEKMLPQ---PPKISEGWDEPTDAALGPPEPCPALEIKMLQAQSPSLPA	223
Qy	209	-----PDFKMWINPCQAVIFDSDPAPKDTGAAA-----LEM-MSQAMIRGMBEGNOF	257
Db	224	PGPPEEBEPIPEPQPTIQASSLPPPDQSARLMAWILHRLFEMALPQPVIIRKGGQEESD-	282
Qy	258	VAPLPVE---ETLKKKRDDEEMDVPDDVDYKIAREY--NNNVKNKASKGYEENYVF	312
Db	283	-----APVTCDVQITISILPGQEE-----SHLILEVDPHW-----	313
Qy	313	FIFREGDGYVYNLETRVRLSKRAKAGVOSGTNALLVVKH-----RDMNKEKELEAQEA	366
Db	314	-----EEDHEOGESTSTSPRISE--AAPADEKGB-----VVEQTPRELPRIQEKED	364
Qy	367	RKAQLENHPEPEEBEHEEMETEKBAAGSDEBKSGSSSEKESGP-DEHSGSESEEREGR	425
Db	365	EDGESEEEGRKEEBEGEKEEBE--GREKEEKEGKEEBEGREKEEBEGEKEDEEGRE	423
Qy	426	DEASDKSGSEDESSDEARAARDKEEIFGSDADSDADSD-----DEDRG	472
Db	424	KEEEGRGKEEESGEGKEEBEGKEVEGREGDEEBEGQDHSVLLDSYLVPOSEEDQS	483
Qy	473	QAQGGSDNSGSGGGQGRSHRSRASPPPGSGHSAQEDGSE-AAASISGSE	525
Db	484	E-----ESETQDQSEVGGAQTQGEVGAQAL---SESETQDQSEVGGAQDQSE	529

RESULT 8

C89824

hypoetical protein sdrC [imported] - Staphylococcus aureus (strain N315)

C/Species: Staphylococcus aureus

C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C/Accession: C89824

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ocu-
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.

A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: C89824

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-953 <R>

A;Cross-references: UNIPROT:Q99W48; GB:BA000018; PID:gl3700453; FIDN:BA841750.1; GSPDB:
A;Experimental source: strain N315

C;Genetics:

A;Gene: sdrC

Qy	35	YCNSLPIDPDP---KFITVPF--DQNRFOYKATSLK-----QHKHDLITPEP-LGVITID	85
Db	375	FVTNLGYFNPDAKNFKIYEVDQNFQVDSFTPDTSKLDVDTQGFVDIYSNDNKATIVD	434
Qy	86	LINPDYRIDPNVLLDPADEKLEEBEIQAPTSKRSQQHAKVVPMMRKTEY-ISTEFNRY	144
Db	435	LLNGOS-----SSDKYIIQOAVYDPNS--STDNGKI-----DYTLETONGKS	475

Query Match 8.5%; Score 234; DB 2; Length 953;

Best Local Similarity 23.0%; Pred.No. 9.9e-09;

Matches 131; Conservative 81; Mismatches 216; Indels 142; Gaps 28;

[illegible]

RESULT 9
D89824
hypothetical protein sdrD [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: D89824
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Ui, Y.; Kohayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D89824
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1385 <R>
A:Cross-references: UNIPROT:Q99W47; GB:BA000018; PID:gl3700454; PID:BAB41751.1; GSPDB:G
A:Experimental source: strain N315
C:Genetics:
A:Gene: sdrD

Query March	813;	Score	230.5;	DB	2;	Length	1385;
Best Local Similarity	20.44;	Fred. No.	0.00023;				
Matches 116;	Conservative	89;	Mismatches	236;	Indels	127;	Gaps
							20;
Qy	27	SGVVCRVKYCNSLR	II	PPDKPFI	YTPDONR	FVQYKATSL	EQKHQHLLTEPDLGV
		:		:			IDL 86
Db	772	TGVI	-----	NGA	NMTLDSG	--	YKTPKYNLGNVY
							WEDTNKDGKQDSTEKGISG
							VTVTL 823
Qy	87	INP	-----	-----	TYRID	-----	PNVLLDPADEKLEER
					:		IQAP 115
Db	824	KNENGEVLQTTK	IK	DKGKYQ	FTGL	ENGYKVE	FETPSGYTPTQVGS
					:		GTDSG-IDSNCTST 882
Qy	116	TSSKRSQ	HA	KV-----	VPMNRKTEY	ISTE	PNRYGISENEKPEVKIGVSV
					:		QQFTSEIYK 170
Db	883	TGVI	IKDKND	NTIDSG	FTKYPT	NLGDY	WEDTNKNGVQDDEKGISG
					:		VTV-----TLK 934
Qy	171	DRDSQITAE	IKETFF	DAQKS	ISQ	-----	HYSKPRVTPEVMPFPDF
					:		KWMINPCAQVIFDS 226

[illegible]

[illegible]

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 13:35:20 ; Search time 167 Seconds
(without alignments)
1229.759 Million cell updates/sec

Title: US-10-721-553-2
Perfect score: 2764
Sequence: 1 MAPTQTQQRGHRPNH.....QEDGSEAAASDSEADSDSD 531

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:.*
1: Geneseq_19808:.*
2: Geneseq_19908:.*
3: Geneseq_20008:.*
4: Geneseq_20018:.*
5: Geneseq_20028:.*
6: Geneseq_20038:.*
7: Geneseq_20039:.*
8: Geneseq_20048:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2764	100.0	531	2	AAY42226 Human pan
2	2764	100.0	531	7	ADD18712 Human dis
3	2764	100.0	531	8	ADO58688 Human reg
4	2764	100.0	531	8	ABM82102 Tumour-as
5	2744	99.3	531	4	AAB93517 Human pro
6	2658.5	96.2	552	4	ABG19682 Novel hum
7	2464	89.1	473	3	AAB56316 Human sec
8	1244.5	45.0	538	4	ABBS59163 Drosophil
9	622	22.5	133	4	ABG19681 Novel hum
10	595	21.5	115	3	AAG03326 Human sec
11	283	10.2	475	4	ABG19412 Human hum
12	283	10.2	475	8	ADS12265 Human the
13	253	9.2	445	6	ABRS3245 Protein s
14	253	9.2	445	7	ADK63670 Disease t
15	237.5	8.6	1633	6	ABU42513 Protein e
16	234.5	8.5	1802	3	AAY83170 Cell wall
17	234.5	8.5	1802	3	AAY70119 Staph. ep
18	234	8.5	952	6	ABU16533 Protein e
19	233.5	8.4	930	2	AAY08641 S. aureus
20	233.5	8.4	947	6	ABJ18940 Pathogen
21	233	8.4	932	4	AAY36845 Staphyloc
22	233	8.4	932	4	AAY34082 Staphyloc
23	232.5	8.4	995	6	ABM72437 Staphyloc
24	230.5	8.3	1385	6	ABU16400 Protein e
25	230	8.3	877	6	ABU42504 Protein e

26	229	8.3	1920	6	ABU43489 Protein e
27	227	8.2	428	5	ABG93245 C. albica
28	227	8.2	567	4	AAE13147 Human ret
29	226	8.2	743	6	ADA89690 Staphyloc
30	226	8.2	773	6	ADA89539 Staphyloc
31	226	8.2	877	6	ABM72702 Staphyloc
32	226	8.2	913	6	ABJ18917 Pathogen
33	225.5	8.2	927	6	ABM72221 Staphyloc
34	225.5	8.2	933	3	AAY58435 Staphyloc
35	225.5	8.2	933	4	AAB69508 Staphyloc
36	225.5	8.2	933	6	ABJ18947 Pathogen
37	225.5	8.2	936	2	AAB89801 Staphyloc
38	224.5	8.1	194	4	ABG11265 Novel hum
39	224	8.1	265	5	ABG32640 Staphyloc
40	223.5	8.1	1315	2	AAY08642 S. aureus
41	223.5	8.1	1315	6	ABJ18969 Pathogen
42	222	8.0	918	2	AAY08640 S. aureus
43	221.5	8.0	1132	2	AAR97866 Chicken l
44	221.5	8.0	1349	4	AAY37544 Staphyloc
45	221.5	8.0	1349	4	AAU34402 Staphyloc

ALIGNMENTS

RESULT 1
AAY42226
ID AAY42226 standard; protein; 531 AA.
XX
AC AAY42226;
XX
DT 20-DEC-1999 (first entry)
XX
DE Human pancreatic differentiation 2 protein sequence.
XX
KW Human; PD2; cancer; regulation; differentiation; neoplastic; therapy;
KW pancreatic differentiation 2; diagnosis; pancreatic adenocarcinoma;
KW phosphoprotein.
XX
OS Homo sapiens.
XX
FN WO9950408-A1.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-US006633.
XX
PR 27-MAR-1998; 98US-0079649P.
XX
PA (UYNE-) UNIV NEBRASKA.
XX
PI Batra SK, Hollingsworth MA;
XX
DR WPI; 1999-591317/50.
XX
N-PSDB; AAZ25433.
XX
PT New phosphoprotein useful as targets for therapy of pancreatic
adenocarcinomas.
XX
PS Claim 7; Fig 2; 97pp; English.
XX
CC The present sequence is the human pancreatic differentiation 2 (PD2)
protein, which comprises an amino terminal helix-loop-helix domain and a
centrally localised nuclear transporter signal and nucleotide binding
site. The PD2 nucleotide sequence and a transformed host cell are useful
for screening a test compounds for PD2 modulating activity indicated by
an alteration in the phosphorylation of status of PD2. The host cells are
assessed for altered expression of pancreatic differentiation markers
(MUC-1 or carbonic anhydrase), and modulating activity is correlated with
an alteration in cellular morphology. The PD gene and protein represent
valuable targets in the differential diagnosis and therapy of pancreatic
adenocarcinomas
XX

AD058688
ID AD058688 standard; protein; 531 AA.
XX
AC AD058688;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human regulatory molecule HRM-9.
XX
KW cytostatic; immunomodulator; agonist; antagonist; gene therapy;
KW human regulatory molecule; HRM; disease development; cell proliferation;
KW immune response; cancer.
XX
OS Homo sapiens.
XX
PN US2002058264-A1.
XX
PD 16-MAY-2002.
XX
PF 26-SEP-2001; 2001US-00840787.
XX
PR 23-SEP-1997; 97US-00933750.
PR 20-JAN-1999; 99US-00234613.
PR 03-MAR-2000; 2000US-00518665.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Hillman JL, Bandman O, Shah P, Au-Young J, Yue H;
PI Guegler KJ, Corley NC;
XX
XX WPI; 2004-459763/43.
DR N-PSDB; ADO58737.
XX
XX New human regulatory molecules, useful in the diagnosis and treatment of
PT cancer and immune disorders.
XX
XX Claim 1; SEQ ID NO 9; 105pp; English.
XX
XX The invention describes human regulatory molecules (HRM) (I) selected
CC from a group comprising the fully defined amino acid sequences of SEQ ID
CC NOS: 1-49. Also described are: an isolated polynucleotide (II) comprising
CC a nucleic acid sequence encoding (I) or the complement of the
CC polynucleotide (SEQ ID NOS:50-98); a composition comprising (II) and a
CC reporter molecule; an expression vector containing (II); a host cell
CC containing the vector; detecting (M1) expression of a nucleic acid in a
CC sample; screening (M2) a plurality of molecules to identify a ligand;
CC diagnosing (M3) a disease associated with gene expression in a sample
CC containing nucleic acids; a composition comprising (I) and a
CC pharmaceutical carrier or a labeling moiety; screening (M4) a plurality
CC of molecules to identify a ligand; preparation and purification of
CC antibodies; an antibody which specifically binds to (I); and detecting
CC protein expression in a sample. The new human regulatory protein
CC molecules which are expressed during disease development and the
CC polynucleotides which encode them satisfies a need in the art by
CC providing compositions which are useful in the diagnosis and treatment of
CC diseases associated with cell proliferation, particularly immune
CC responses and cancers. This is the amino acid sequence of a human
CC regulatory molecule.
XX
SQ Sequence 531 AA;

Query Match 100.0%; Score 2764; DB 8; Length 531;
Best Local Similarity 100.0%; Pred. No. 1.3e-196;
Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPTIQQAQEDGHRPNSHRTLPSRSGVCRVKYCNLSLPIDPDPKFIITFPDQNEFVQ 60
DB 1 MAPTIQQAQEDGHRPNSHRTLPSRSGVCRVKYCNLSLPIDPDPKFIITFPDQNEFVQ 60
QY 61 YKATSLKQKHDLITLTPDLGVTTDLINPDYRIDPNVLLDPADEKLEEEIQAPTSKR 120
DB 61 YKATSLKQKHDLITLTPDLGVTTDLINPDYRIDPNVLLDPADEKLEEEIQAPTSKR 120

QY 121 SQOHAKVPMRKTEYISTEFNRYGISNEKPEVKIGSVKQOFTETEEIYKDRDSQITAE 180
DB 121 SQOHAKVPMRKTEYISTEFNRYGISNEKPEVKIGSVKQOFTETEEIYKDRDSQITAE 180
QY 181 KTFEDAQKSIHQHYSKPRVTVPVEVMFVPDFKQWNPACQVIFDSDPAPKD7SGAAALEM 240
DB 181 KTFEDAQKSIHQHYSKPRVTVPVEVMFVPDFKQWNPACQVIFDSDPAPKD7SGAAALEM 240
QY 241 MSQAMIRGMMDSEGNQFVAYFLPVEETLKQRKDDQEEEMDYAPDDVYDYKIAREYNVVK 300
DB 241 MSQAMIRGMMDSEGNQFVAYFLPVEETLKQRKDDQEEEMDYAPDDVYDYKIAREYNVVK 300
QY 301 NKASKGYEENYFPIFREGDGVYVYNELETRVRLSKRAKAGVQSGTALLVVHHRDMNEKE 360
DB 301 NKASKGYEENYFPIFREGDGVYVYNELETRVRLSKRAKAGVQSGTALLVVHHRDMNEKE 360
QY 361 LEAQEARKAQLENHPEPEEEEEEMETEKEAGGSEDEEKEGSSSEKESGSDHSGSSESR 420
DB 361 LEAQEARKAQLENHPEPEEEEEEMETEKEAGGSEDEEKEGSSSEKESGSDHSGSSESR 420
QY 421 EECGRDEASDKSGGDESEDEARAAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480
DB 421 EECGRDEASDKSGGDESEDEARAAARDKEEIFGSDADSEDDADSDDEDRGQAQGGSDN 480
QY 481 DSDSGSGGGQSRSHRSRSGASPPSGSEHSAQDGSSEAAASDSSEADSDSD 531
DB 481 DSDSGSGGGQSRSHRSRSGASPPSGSEHSAQDGSSEAAASDSSEADSDSD 531

RESULT 4
ABM82102
ID ABM82102 standard; protein; 531 AA.
XX
AC ABM82102;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) polypeptide PRO83014, SEQ:5424.
XX
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic.
OS Homo sapiens.
XX
XX WO2004030615-A2.
XX
XX 15-APR-2004.
XX
XX 29-SEP-2003; 2003WO-US028547.
XX
XX 02-OCT-2002; 2002US-0414971P.
XX
XX (GETH) GENENTECH INC.
XX
XX Wu TD, Zhang Z, Zhou Y;
XX
XX WPI; 2004-347921/32.
DR N-PSDB; ACN40565.
XX
XX New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
XX Claim 12; SEQ ID NO 5424; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are

overexpressed in cancer tissues compared to normal tissues, and may thus serve as effective targets for the diagnosis and treatment of cancer in mammals. The invention also relates to nucleic acid and polypeptide sequences at least 80% identical to the TAT nucleic acids and polypeptides; expression vectors and host cells comprising a TAT nucleic acid; an antibody specific for a TAT polypeptide; a peptide or organic molecule which binds to a TAT polypeptide; fusion proteins comprising a TAT polypeptide; and methods and compositions for the treatment or diagnosis of cancer in mammals. TAT polypeptides, nucleic acids, antibodies, antagonists, binding molecules and compositions are useful for diagnosing or treating a cell proliferative disorder associated with increased TAT expression, particularly cancers such as breast cancer, colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder cancer, pancreatic cancer, cervical cancer, cancers of the central nervous system, melanoma and leukaemia. TAT nucleic acids may further be used as hybridisation probes, in chromosome and gene mapping, in chromosome identification and in gene therapy. The present sequence represents a TAT polypeptide of the invention

Sequence 531 AA;

Query Match 100.0%; Score 2764; DB 8; Length 531;
 Best Local Similarity 100.0%; Pred. No. 1.3e-196;
 Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPIQTQAQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRVQ 60
 Db 1 MAPIQTQAQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRVQ 60

Qy 61 YKATSLKQHKHDLTPDPLGVTTIDLPNVTYRIDPNVLLDPADKLEETQAPTSSKR 120
 Db 61 YKATSLKQHKHDLTPDPLGVTTIDLPNVTYRIDPNVLLDPADKLEETQAPTSSKR 120

Qy 121 SQQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITAE 180
 Db 121 SQQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITAE 180

Qy 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDFKQWNPVCAQVIFDSDPAPKPTSGAAALEM 240
 Db 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDFKQWNPVCAQVIFDSDPAPKPTSGAAALEM 240

Qy 241 MSQAMTGMWDESNQVAVFLPVEETLKRRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300
 Db 241 MSQAMTGMWDESNQVAVFLPVEETLKRRKRDQEEEMDYAPDDVYDYKIAREYNWVK 300

Qy 301 NKASKGYEENYFTIFREGDGVYVNELETRVRLSKRAKAGVQSGTGNALLVVKHRDMNEKE 360
 Db 301 NKASKGYEENYFTIFREGDGVYVNELETRVRLSKRAKAGVQSGTGNALLVVKHRDMNEKE 360

Qy 361 LEAQEARKAQLNHEPEEEEEEMETEKEAGGSDEEKEKGSSEKEGSEDEHSGSESER 420
 Db 361 LEAQEARKAQLNHEPEEEEEEMETEKEAGGSDEEKEKGSSEKEGSEDEHSGSESER 420

Qy 421 EEGDRDASPKSGSGEDESSEDARAKDKEEITFGSDADSEDDADSDDEDRGQAQGGSDN 480
 Db 421 EEGDRDASPKSGSGEDESSEDARAKDKEEITFGSDADSEDDADSDDEDRGQAQGGSDN 480

Qy 481 DSDSGSGGGQGRSHRSRSPFPSSGSEHSAQEDGSEAAASDSEADSDSD 531
 Db 481 DSDSGSGGGQGRSHRSRSPFPSSGSEHSAQEDGSEAAASDSEADSDSD 531

RESULT 5

AAB93517

ID AAB93517 standard; protein; 531 AA.

XX

AC AAB93517;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:12853.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 12853; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 531 AA;

XX Query Match 99.3%; Score 2744; DB 4; Length 531;

XX Best Local Similarity 99.4%; Pred. No. 3.9e-195;

XX Matches 528; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MAPIQTQAQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRVQ 60

Db 1 MAPIQTQAQREDGHRPNHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRVQ 60

Qy 61 YKATSLKQHKHDLTPDPLGVTTIDLPNVTYRIDPNVLLDPADKLEETQAPTSSKR 120

Db 61 YKATSLKQHKHDLTPDPLGVTTIDLPNVTYRIDPNVLLDPADKLEETQAPTSSKR 120

Qy 121 SQQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITAE 180

Db 121 SQQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITAE 180

Qy 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDFKQWNPVCAQVIFDSDPAPKPTSGAAALEM 240

Db 181 KTFEDAQKSIQSHYSKPRVTPVEVMPVFPDFKQWNPVCAQVIFDSDPAPKPTSGAAALEM 240

QY 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDYKILAREYNNVK 300
DB 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDYKILAREYNNVK 300
QY 301 NKASKGYENYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTGNALLVVKHRDMNEKE 360
DB 301 NKASKGYENYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTGNALLVVKHRDMNEKE 360
QY 361 LEAQEARKAOLENHEPEEEEEEMETEKEAGGSDDEOEKSSSEKEGSEDEHSGSSESER 420
DB 361 LEAQEARKAOLENHEPEEEEEEMETEKEAGGSDDEOEKSSSEKEGSEDEHSGSSESER 420
QY 421 EGRDRDEASDKSGGSEDESSDEAARAADKBEIFGSDADSDDDADSDDEDRGQAGGSDN 480
DB 421 EGRDRDEASDKSGGSEDESSDEAARAADKBEIFGSDADSDDDADSDDEDRGQAGGSDN 480
QY 481 DSDSGSGGGQRRSHRSASPPFSGSHSAQEDGSEAAASDSEADSDSD 531
DB 481 DSDSGSGGGQRRSHRSASPPFSGSHSAQEDGSEAAASDSEADSDSD 531

RESULT 6
ABG19682
ID ABG19682 standard; protein; 553 AA.
XX AC ABG19682;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #19673.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US0008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX FA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-FSDB; AAS83869.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PS biodiversity.
XX PS Claim 20; SEQ ID NO 50041; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 553 AA;
Query Match 96.2%; Score 2658.5; DB 4; Length 553;
Best Local Similarity 95.0%; Pred. No. 9.2e-189;
Matches 515; Conservative 6; Mismatches 10; Indels 11; Gaps 1;
QY 1 MAPIOTQAOERDGHHPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFPDQNFVQ 60
DB 12 MAPIOTQAOERDGHHPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFIYFPDQNFVQ 71
QY 61 YKATSLKQKHDLLETPDLGVITIDLINPTYRIDPNVLLDPADEKLEBEIQAPTSSKR 120
DB 72 YKATSLKQKHDLLETPDLGVITIDLINPTYRIDPNVLLDPADEKLEBEIQAPTSSKR 131
QY 121 SQQHAVVPMRKTEYISTEFNRYGISTNEKPEVKIGVSVKQOFTTEEIIYKDEDSQITAE 180
DB 132 SQQHAVVPMRKTEYISTEFNRYGISTNEKPEVKIGVSVKQOFTTEEIIYKDEDSQITAE 191
QY 181 KTFEDAQS-----ISQHSKPRVTPVEVMVFPDFKWINPQCAQVIFDSPPAP 229
DB 192 KTFEDAQSVIEGLGWGEARISQHSKPRVTPVEVMVFPDFKWINPQCAQVIFDSPPAP 251
QY 230 KDTSGAAALEMMSQAMIRGMDDEGNQFVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDY 289
DB 252 KDTSGAAALEMMSQAMIRGMDDEGNQFVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDY 311
QY 290 KIAREYNWNVKNKASKGYEENYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTNALL 349
DB 312 KIAREYNWNVKNKASKGYEENYFFIFREGDGVYVNELETRVRLSKRAKAGVQSGTNALL 371
QY 350 VVHRDNEKELEAQEARKAOLENHEPEEEEEEMETEKEAGGSDDEOEKSSSEKEGS 409
DB 372 VVHRDNEKELEAQEARKAOLENHEPEEEEEEMETEKEAGGSDDEOEKSSSEKEGS 431
QY 410 EDEHSGSESEREGDRDEASDKSGGSEDESSDEAARAADKBEIFGSDADSDDEADSDDE 469
DB 432 EDEHSGSESEREGDRDEASDKSGGSDSDYKARAARDKKEIFGSDADSDDEADSDDE 491
QY 470 DRGQAQGGSDNDSGSGGGQRRSHRSASPPFSGSHSAQEDGSEAAASDSEADSDSD 529
DB 492 DRGQAQGGSDNDSGSGGGQRRSHRSASPPFSGSHSAQEDGSEAAASDSEADSDSD 551
QY 530 SD 531
DB 552 SD 553

RESULT 7
AAB56316
ID AAB56316 standard; protein; 473 AA.
XX AC AAB56316;
XX DT 13-MAR-2001 (first entry)
XX DE Human secreted protein sequence encoded by gene 106 SEQ ID NO:410.
XX KW Human; secreted protein; diagnosis; immunosuppressive; antithrptic;
XX KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
XX KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
XX KW fungicide; ophthalmological; gene therapy; pathological condition;
XX KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
XX KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
XX KW cerebrovascular disorder; angiogenesis; nervous system disorder;

KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; skin aging; food additive; preservative.
 XX Homo sapiens.
 XX WO200070042-A1.
 XX 23-NOV-2000.
 XX 11-MAY-2000; 2000WO-USP12788.
 XX 13-MAY-1999; 99US-0134068P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
 PI Duan RD, Florence KA, Soppet DR;
 XX WPI; 2000-679828/66.
 XX Isolated nucleic acid molecule encoding a human secreted protein is used
 PT in preventing, treating or ameliorating a medical condition.
 XX Disclosure; Page 1041-1042; 1065pp; English.
 XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
 CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
 CC proteins have activities based on the tissues and cells the genes are
 CC expressed in. Examples of activities include: immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
 CC virucide; fungicide; and ophthalmological. The human secreted
 CC polynucleotides and proteins can be used to prevent, treat or ameliorate
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
 CC dogs, chickens or sheep. They are also used in diagnosing a pathological
 CC condition or susceptibility to a pathological condition. Disorders which
 CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
 CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
 CC fungi and ocular disorders e.g. corneal infection. The proteins can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The proteins can also be used as a
 CC food additive or preservative to increase or decrease storage
 CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
 CC in the exemplification of the present invention
 XX SQ Sequence 473 AA;
 Query Match 8p.1%; Score 2464; DB 3; Length 473;
 Best Local Similarity 9p.8%; Pred. No. 2.1e-174;
 Matches 472; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 MAPTIQTAQREDGHRNSHRTLPERSGVGVCRVKYCNLSLDPIDFPDPKFTYPPDQNRVQ 60
 Db 1 MAPTIQTAQREDGHRNSHRTLPERSGVGVCRVKYCNLSLDPIDFPDPKFTYPPDQNRVQ 60
 Qy 61 YKATSLKQHKHDLTTPDGLGVTIDLPNTYRIDPNVLDPADEKLLBEEIQAPTSKKR 120
 Db 61 YKATSLKQHKHDLTTPDGLGVTIDLPNTYRIDPNVLDPADEKLLBEEIQAPTSKKR 120
 Qy 121 SQQKAVVPMWRKTEYISTEFNRXGISNEKPEVKIGVSVKQFTBEEIYKDRDSQITAE 180
 Db 121 SQQKAVVPMWRKTEYISTEFNRXGISNEKPEVKIGVSVKQFTBEEIYKDRDSQITAE 180
 Qy 181 KTFEDAQKSIQSKYKPRVTFVEMVFPDFKWINPACQVIFDSDPAPKDTSGAAALEM 240
 Db 181 KTFEDAQKSIQSKYKPRVTFVEMVFPDFKWINPACQVIFDSDPAPKDTSGAAALEM 240
 Qy 241 MSQAMIRGMDEEGNQVAFLPVSEETLKKRRDQDEEMDYAPDDVDYDKLAEYNNVK 300

Db 241 MSQAMIRGMDEEGNQVAFLPVSEETLKKRRDQDEEMDYAPDDVDYDKLAEYNNVK 300
 Qy 301 NKASKGYENYFFIFREGDGVYVYNELETRVRLSKRAKAGVQSGTNNALLVMKHRDMNEKE 360
 Db 301 NKASKGYENYFFIFREGDGVYVYNELETRVRLSKRAKAGVQSGTNNALLVMKHRDMNEKE 360
 Qy 361 LEAQARKAQLNHEPHEEEEEEMETEEKACGSGDEEQKSGSSEKESGSEDEHSGSESER 420
 Db 361 LEAQARKAQLNHEPHEEEEEEMETEEKACGSGDEEQKSGSSEKESGSEDEHSGSESER 420
 Qy 421 EBGDRDEASDKSGSGDESDSEARAARKEIFGSDADSEDDADSDDDDRGQ 473
 Db 421 EBGDRDEASDKSGSGDESDSEARAARKEIFGSDADSEDDADSDDDDRGQ 473
 RESULT 8
 ABB59163
 ID ABB59163 standard; protein; 538 AA.
 XX ABB59163;
 AC ABB59163;
 XX 26-MAR-2002 (first entry)
 DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 4281.
 DE Drosophila melanogaster polypeptide SEQ ID NO 4281.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL03266.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX Disclosure; SEQ ID NO 4281; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
 CC sequences (ABLI01840-ABLI6175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 538 AA;
 Query Match 45.0%; Score 1244.5; DB 4; Length 538;
 Best Local Similarity 50.0%; Pred. No. 9.3e-84;
 Matches 271; Conservative 66; Mismatches 172; Indels 33; Gaps 11;
 Qy 1 MAPTIQTAQREDGHRNSHRTLPERSGVGVCRVKYCNLSLDPIDFPDPKFTYPPDQNRVQ 60
 Db 1 MPPTINNSAVNSAEK-RPQRTERKSEIICRVKGNLDPIDFPDLKFLQKFPDHRFVQ 59

PI	Dumas Milne Edwards J,	Duclert A, Giordano J;
XX	WPI; 2000-500381/45.	
DR	N-PSDB; AAC03332.	
XX		
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
PT	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for	
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures.	
XX		
PS	Claim 13; SEQ ID NO 7407; 71pp + Sequence Listing; English.	
XX		
CC	The present sequence is a polypeptide encoded by one of a large number of	
CC	5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were	
CC	prepared from total human RNAs or polyA+ RNAs derived from 30 different	
CC	tissues. EST sequences usually correspond mainly to the 3' untranslated	
CC	region (UTR) of the mRNA because they are often obtained from oligo-dT	
CC	primed cDNA libraries. Such ESTs are not well suited for isolating cDNA	
CC	sequences derived from the 5' ends of mRNAs and even in those cases where	
CC	longer cDNA sequences have been obtained, the full 5' UTR is rarely	
CC	included. 5' ESTs are derived from mRNAs with intact 5' ends and can	
CC	therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs	
CC	are also used in diagnostic, forensic, gene therapy and chromosome	
CC	mapping procedures. They are used to obtain upstream regulatory sequences	
CC	and to design expression and secretion vectors	
XX		
SQ	Sequence 115 AA;	
	Query Match 21.5%; Score 595; DB 3; Length 115;	
	Best Local Similarity 95.5%; Pred. No. 2,1e-36;	
	Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0	
Qy	1 MAPTIQTQAQREDGHRPNSHRTLPERSGVCRVKYCNLSLPDIPDPKFTITYPFDQNRVQ 60	
Db	1 MAPTIQTQAQREDGHRPNSHRTLPKXSGVCRVKYCNLSLPDIPDPKFTITYPFDQNRVQ 60	
Qy	61 YKATSLKQHKHDLITTEPDGLGVTIDLINPDTRYIDPNVLLDPADEKLEESIQAP 115	
Db	61 YKATSLKQHKHDLITTEPDGLGVTIDLINPDTRYIDPNVLLDPADEKLEESIQXP 115	
RESULT 11		
ABG19412		
ID	ABG19412 standard; protein; 475 AA.	
XX		
AC	ABG19412;	
XX		
DT	13-FEB-2002 (first entry)	
XX		
DE	Novel human diagnostic protein #19403.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001WO-US008631.	
XX		
PR	31-MAR-2000; 2000US-00540217.	
PR	23-AUG-2000; 2000US-00549167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Drmanac RT, Liu C, Tang YT;	
XX		
DR	WPI; 2001-639362/73.	
DR	N-PSDB; AAS83599.	
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	
PT	diagnostics, forensic, gene mapping, identification of mutations	

PI Tang YT, Asundi V, Ren F, Zhang J, Wehrman T, Wang Z, Ma Y; Zhou P;
PI Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;
XX WPI: 2004-668857/65.
DR N-PSDB; ADS11667.
XX New polynucleotide, useful in preparing a composition for diagnosing or
PT treating inflammatory, neurodegenerative or stem cell disorders, e.g.,
PT aplastic anemia or cancer for promoting wound healing.
XX
XX Example 2; SEQ ID NO 2502; 718pp; English.
XX The invention relates to a novel isolated polynucleotide and the encoded
CC polypeptide. The molecules of the invention demonstrate antiinflammatory,
CC neuroprotective, antianemic, cytostatic and vulnerary activities and may
CC be useful in preparing a composition for diagnosing or treating
CC inflammatory, haematopoietic, immune, neurodegenerative or stem cell
CC disorders, such as aplastic anaemia or cancer, as well as for promoting
CC wound healing. The molecules may also be utilised during gene therapy
CC procedures. The current sequence is that of a human therapeutic contig
XX protein of the invention.
XX
SQ Sequence 475 AA;

Query Match 10.2%; Score 283; DB 8; Length 475;
Best Local Similarity 87.7%; Pred. No. 2e-12;
Matches 57; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 328 TTVRLSKRAKAGVCSGTNALLVVKHRDWNKELEAQAARKAQLNHPPEEEEMETE 387
DB 3 SRVLSKRAKAGVCSGTNALLVVKHRDWNKELEAQAARKAQLNHPPEEEEMETEIRQP 62

QY 388 EKEAG 392
DB 63 RKLIG 67

RESULT 13
ID ABR53245
XX ABR53245 standard; protein; 445 AA.
XX
AC ABR53245;
XX
DT 20-JUN-2003 (first entry)
DE Protein sequence #SEQ ID 1355.
XX
XX Multiprotein complex; eukaryote; drug target; diagnosis.
XX
XX Saccharomyces cerevisiae.
XX
XX EP1258494-A1.
XX
PD 20-NOV-2002.
XX
PF 20-DEC-2001; 2001EP-00130253.
XX
PR 15-MAY-2001; 2001EP-00111774.
XX
PA (CELL-) CELLZONE AG.
XX
PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
PI Marzioch M, Schultz JD, Superti-Furga GD;
XX
DR WPI: 2003-250078/25.
DR N-PSDB; ACC61287.
XX
XX New isolated protein complexes useful for diagnosing a disease or
PT disorder, or as a target for an active agent of a pharmaceutical,
PT preferably a drug target in the treatment or prevention of disease or
PT disorder.
XX
PS Disclosure; SEQ ID NO 1355; 17pp + Sequence Listing; English.

XX The invention relates to multiprotein complexes from eukaryotes. Proteins
CC of the invention and DNA sequences encoding them are given in records
CC ABR52568-ABR53903 and ACC610-ACC61944 respectively. The complexes are
CC obtainable by using a protein as a bait and isolating the set of proteins
CC which is attached thereto from cells. Such protein complexes may comprise
CC up to 30 distinct proteins. Protein complexes of the invention are useful
CC for diagnosing a disease or disorder, or as a target for an active agent
CC of a pharmaceutical, preferably a drug target in the treatment or
CC prevention of a disease or disorder. Note: The sequence data for this
CC patent is not represented in the printed specification, but is based on
CC sequence information supplied by the European Patent Office. The complete
CC document is available on CD-ROM
XX
SQ Sequence 445 AA;

Query Match 9.2%; Score 253; DB 6; Length 445;
Best Local Similarity 22.1%; Pred. No. 3.2e-10;
Matches 112; Conservative 96; Mismatches 184; Indels 114; Gaps 21;

QY 23 LPERSGVVCVKYCNLSLPDIPDPKFTYTP-----FDQNFVQYKATSLKQHK 71
DB 1 MSKKQEIYIPIKYQNSLPVLPQLPPLLVYPESPETNADSSQLINSLYIKTNVTNLIQ-- 58

QY 72 HDLLTEBDLGVITDLI-----NPDYRIDPNVLLDPADEKLEEEIQAPTSRS 121
DB 59 -----DEDLGMPVDLMKFPGLLNKLDKSLYGFDP-NVKLDKDDILLRD---PRIDRLT 108

QY 122 QOHAKVPMVRKTEYISTEFNRYGISNEKPEVKIGSVKQOFTEEIYKDRDSQ----- 175
DB 109 KTDISKVTFLRRTEYVSVNTIAAHDNTSLKRRRL-----DDGSDDENLDV 154

QY 176 ---ITAIEKTFEDAQKSIQHYSKPRVTPVEVMVFPDFKMNINPCAQVTPSDPAPKDT 232
DB 155 NHIISRVEGTENKTDK--WQHPVKKGVMVKCKWLLPD---TASMDQVYF-----ILKF 203

QY 233 SGAAALEMMSQAMTR-CM---MDEEGNQFVAYFLPVEETLKKRCDDEEWDYAPDDVVD 288
DB 204 MGSASLDTKSKKSLNTGIFRPVELEDEWISMYATDHSKSAILENELEKGMDEDDSHSHE 263

QY 289 ---YKIAREYNMNVKNKASKGYENYFFIFREGDGV-YTNELETRVRLSKRAKAGVQSG 344
DB 264 GKIYKPRIRDYDMKQVAEKPMTE-LAIRLNDKOGIAYYKPLRSKIELRRVRVNDIIP- 321

QY 345 TNALLVVVKH-----RDMNEKELEAQAARKAQLN-----HPPEEEEMETE 389
DB 322 ----LVKEHDIDQLNVTLRNPSTKEANIRDKLRMKFDPINPATVDEEDDEDEQEDVYKK 377

QY 390 FAGSDEEQEGSSSEKSGSEDEHSGSESEEEGDRDEASDKSGSDESEDEARAAARD 449
DB 378 ESEG--DSKTEGSEQEGENKDEIKOEKNEQ-----DEENKQDENRAADT 422

QY 450 KEEIFGSDADSDDDADSDDDDRGQAQ 475
DB 423 PET---SDAVHTQKPEEKETLQEE 445

RESULT 14
ID ADK63670
XX ADK63670 standard; protein; 445 AA.
XX
AC ADK63670;
XX
DT 06-MAY-2004 (first entry)
XX
DE Disease treating protein complex-derived protein #818.
XX
KW protein complex; drug target; diagnosis.
XX
OS Unidentified.
XX
PN EP1338608-A2.
XX

27-AUG-2003.
20-DEC-2002; 2002EP-00102902.
20-DEC-2001; 2001EP-00130253.
(CELL-) CELLZOME AG.
Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;
Michon A, Leutwein C, Rick J;
WPI; 2003-638460/61.
N-PSDB; ADK63671.
New proteins and protein complexes from eukaryotes, useful as targets in drug screening, or in diagnosing or screening for the presence of a disease or disorder, or a predisposition for developing a disease or disorder in a subject.
Disclosure; SEQ ID NO 1635; 13pp; English.
The invention relates to novel protein complexes comprising a first and a second protein, or its derivative, fragment, homologue or variant. The proteins are selected from given protein complexes, which are not defined in the specification. The variants are encoded by nucleic acids that hybridize to the nucleic acids encoding the proteins under low stringency conditions. The protein complexes are useful as targets for an active agent of a pharmaceutical. These protein complexes are particularly useful as drug targets for the treatment or preventing of a disease or disorder. The complexes and methods above are useful in diagnosing or screening for the presence of a disease or disorder or a predisposition for developing a disease or disorder in a subject. These are also useful in screening for a drug for treatment or prevention of a disease or disorder. The molecule that modulates the amount, activity or protein components of the complex is useful for the manufacture of a medicament for the treatment or prevention of a disease or disorder. This sequence corresponds to a protein of the invention. (Note: the sequence data for this patent did not form part of the printed specification but was obtained from the EPO in electronic format).
SQ Sequence 445 AA;
Query Match 9.28; Score 253; DB 7; Length 445;
Best Local Similarity 22.18; Pred. No. 3.2e-10;
Matches 112; Conservative 96; Mismatches 184; Indels 114; Gaps 21;
QY 23 LPERSGVVCVKYCNLSLPDIPDFPKFTYP-----FQNRVQVKATSLKQHK 71
DB 1 MSKQEVIAPIKQNSLPVQLPKLVYSPETNADSSQLNSLYIKTNVNLQQ-- 58
QY 72 HDLLTEPLDGLVTIDLI-----NPDYRIDPNVLLDPADEKLEELIQAPTSQRS 121
DB 59 -----DEDLGMPVLMKPPGLLNKLDKLLYGFD-NVKLDKDRILLRD----PRIDRLT 108
QY 122 QOHAQVVPWVKTEYISTEFNRYGISNEKEPVKGVSKQFTEEEYKDRDSQ----- 175
DB 109 KTDISKVTEFLRTYEVNTIAHDNTSLKRRRL-----DGDSDDENLDV 154
QY 176 ---ITALEKTFEQAQKISQHSYSPRTVPVEMVFPDFKMWINPCAQVIFDSDPAKDT 232
DB 155 NHIISRVEGTFNKTDK--WQHPVKGVKVMKWKDLLFD-----TASMDQVVF----ILKF 203
QY 233 SGAAALEMMSOAMIR-GM---MBEGNQFVAYFLPVEETLKRRDQOEEMDVAPDDVD 288
DB 204 MGSASLTQKSKSLNTGIFRPVELEDEWISWATDKDSAILLENELEKGMDEMDDDSHE 263
QY 289 ---YKIAREYNWKNKASKGYENYFFIFREGDGV--YNELETRVRLSKRAKAGVQS 344
DB 264 GKIKYFKRIIRDYDKQVAKPMTE-LAIRLNDKDGIAIYKPLSKLELRRRRVNDLIK- 321
QY 345 TNALLVVKH-----RDMNEKELEAQBARKAQLN-----HPEEBEEMETEER 389
DB 322 ----LVKEHDIDQLNVTLRNPSTKEANIRDKLRMKFDPINFATVDEDEDEEQPEDVK 377
QY 390 EAGGSEDEQKSGSSEKSGSEHSGSESEREGDRDEADSKSGSGEDSSSEDEARAARD 449
DB 378 ESEG--DSKTEGSEQEGENEKDEEIKQEKENEQ-----DEENKQDENRAADT 422
QY 450 KEEIFGSDADSDSDADSDDDDRGQAQ 475
DB 423 PET---SDAVHTEQKPEEKEKTLQEE 445
RESULT 15
ID ABU42513 standard; protein; 1633 AA.
XX ABU42513;
AC ABU42513;
XX 19-JUN-2003 (first entry)
DT Protein encoded by Prokaryotic essential gene #28040.
DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Staphylococcus epidermidis.
OS WO200277183-A2.
PN 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA46383.
XX New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.
PT Claim 25; SEQ ID NO 70437; 1766pp; English.
PS The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of the strains; or (13) identifying the target of a compound that inhibits proliferation of an organism. The antisense nucleic acids are useful for

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2005, 13:36:10 ; Search time 179 Seconds
(without alignments)
1519.073 Million cell updates/sec

Title: US-10-721-553-2

Perfect score: 2764

Sequence: 1 MAPTIQTQAREDHGRPNH.....QEDGSEAAASDSSSEADSDSD 531

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2764	100.0	531	2 Q9H166	Q9H166 homo sapien
2	2744	99.3	531	2 Q9NUU9	Q9NUU9 homo sapien
3	2718	98.3	535	2 Q8K2T8	Q8K2T8 mus musculus
4	2708	98.0	535	2 Q9J199	Q9J199 mus musculus
5	2531.5	91.6	510	2 Q75239	Q75239 homo sapien
6	2183.5	79.0	520	2 Q6F2Y1	Q6F2Y1 xenopus tro
7	1995	72.2	485	2 Q8N7H5	Q8N7H5 homo sapien
8	1984	71.8	377	2 Q9CS63	Q9CS63 mus musculus
9	1935	70.0	407	2 Q88F51	Q88F51 xenopus lae
10	1244.5	45.0	538	2 Q9VN55	Q9VN55 drosophila
11	1129	40.8	468	2 Q7PXA3	Q7PXA3 anopheles g
12	645	23.3	425	2 Q90783	Q90783 caenorhabdi
13	372.5	13.5	589	2 Q8RW91	Q8RW91 arabidopsis
14	361.5	13.1	593	2 Q9MA04	Q9MA04 arabidopsis
15	358.5	13.0	451	2 Q6ZD92	Q6ZD92 oryza sativ
16	335.5	12.1	547	2 Q9CA82	Q9CA82 arabidopsis
17	286.5	10.4	386	2 Q6C509	Q6C509 yarrowia li
18	280.5	10.1	457	2 Q9US06	Q9US06 schizosacch
19	265.5	9.6	791	2 Q8DGL1	Q8DGL1 figu rubrip
20	257.5	9.3	571	2 Q8MTN7	Q8MTN7 trichinella
21	253	9.2	445	1 PAF1_YEAST	P38351 saccharomyc
22	251.5	9.1	538	2 Q9ET15	Q9ET15 mus musculus
23	250.5	9.1	458	2 Q6BT93	Q6BT93 debaryomyce
24	245.5	8.9	784	2 Q7LZ90	Q7LZ90 torpedo cal
25	245	8.9	1848	2 Q7RG8P	Q7RG8P bos taurus
26	244.5	8.8	934	2 Q9GMD3	Q9GMD3 bos taurus
27	243.5	8.8	438	2 Q6FXJ9	Q6FXJ9 candida gla
28	241	8.7	792	2 Q9YTL7	Q9YTL7 ateline her
29	240	8.7	1394	1 CNG4_BOVIN	Q2B181 bos taurus
30	238.5	8.6	946	2 Q7QC53	Q7QC53 anopheles g
31	238	8.6	613	2 Q6UDM5	Q6UDM5 brachydanio

32	236.5	8.6	1633	2	Q8CMP4	Q8CMP4 staphylococ
33	235	8.5	1451	2	Q8I2D8	Q8I2D8 plasmodium
34	234.5	8.5	1733	2	Q9KI14	Q9KI14 staphylococ
35	234	8.5	498	2	Q8MTN8	Q8MTN8 trichinella
36	234	8.5	953	2	Q9NW48	Q9NW48 staphylococ
37	234	8.5	953	2	Q7A781	Q7A781 staphylococ
38	234	8.5	1323	2	Q9NB35	Q9NB35 plasmodium
39	233.5	8.4	947	2	Q86487	Q86487 staphylococ
40	233	8.4	955	2	Q8NXX7	Q8NXX7 staphylococ
41	232	8.4	957	2	Q6GBS6	Q6GBS6 staphylococ
42	231.5	8.4	873	2	Q6GDH2	Q6GDH2 staphylococ
43	230.5	8.3	843	2	Q6GLM0	Q6GLM0 xenopus lae
44	230.5	8.3	897	2	Q13098	Q13098 xenopus lae
45	230.5	8.3	1385	2	Q99W47	Q99W47 staphylococ

ALIGNMENTS

RESULT 1
Q9H166 PRELIMINARY; PRT; 531 AA.
AC Q9H166; 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein PD2.
GN Name=PD2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Batra S.K., Choudhury A., Keita M., Schmiel B.M., Vanlith M.,
Walter N.A.R., Jorkest J., Sikela J.M., Metzgar R.S.,
Hollingsworth M.A.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle, and Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Udén T.B., Toshyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RA Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ401156; CAC20564.1; -.

01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
RIKEN cDNA 5730511K23.
GN Name=5730511K23Rik;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Mammary tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg E., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.P., Casavant T.L., Schetz T.F.,
RA Brownstein M.J., Udell J.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Mammary tumor;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC029843; AAH29843.1; -;
DR MGD; MGI:1923988; 5730511K23Rik.
DR InterPro; IPR007133; Pfam.
DR Pfam; PF03985; Pfam; 1.
SQ SEQUENCE 535 AA; 60538 MW; 7A5EAB1284988070 CRC64;

Query Match 98.13%; Score 2718; DB 2; Length 535;
Best Local Similarity 98.11%; Pred. No. 7,1e-121;
Matches 525; Conservative 2; Mismatches 4; Indels 4; Gaps 1;
QY 1 MAPTTQQAQREDGHRPNSHRTLPSRGVGVCRVKYCNLSLPDIPDPKPTTYPFDQNRVQ 60
DB 1 MAPTTQQAQREDGHRPNSHRTLPSRGVGVCRVKYCNLSLPDIPDPKPTTYPFDQNRVQ 60
QY 61 YKATSLKQKHDLITLTPDLGVITIDLINPDTYRIDPNVLLPADKLEEEIQAPTSSKR 120
DB 61 YKATSLKQKHDLITLTPDLGVITIDLINPDTYRIDPNVLLPADKLEEEIQAPTSSKR 120
QY 121 SQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTFFTEEYIKRDSQITAE 180
DB 121 SQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTFFTEEYIKRDSQITAE 180
QY 181 KTFEDAQKSIHQYXSKPRVTPVEVMPVFPDFQWNPVCAQVIFDSDPAKPTSGAAALEM 240
DB 181 KTFEDAQKSIHQYXSKPRVTPVEVMPVFPDFQWNPVCAQVIFDSDPAKPTSGAAALEM 240
QY 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRKDOEEEMDYAPDDVDYKIDARENNVVK 300
DB 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRKDOEEEMDYAPDDVDYKIDARENNVVK 300
QY 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRKAQVQSGTNALLVVKHRDMNEKE 360
DB 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRKAQVQSGTNALLVVKHRDMNEKE 360
QY 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDSDDEHQAGGSDN 480
DB 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDSDDEHQAGGSDN 480

361 LEAQARKAQLNHPPEEEEEEMEAEEKKAGGSEDEQEKSGSSEKSGSEDEHSGSESDR 420
QY 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDSDDEHQAGGSDN 480
DB 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDSDDEHQAGGSDN 480
QY 481 DSDSGSNGGQGR-----SRSHRSASPPPSGSEHSAQEDGSEAAAAADSDSEADSDSD 531
DB 481 DSDSGSNGGQGRSRSRQSRSRSRSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 535
RESULT 4
Q9J999 PRELIMINARY; PRT; 535 AA.
ID Q9J999
AC Q9J999
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mus musculus brain cDNA, clone MNCb-6444, similar to Homo sapiens cDNA
FLJ11123, clone PLACE1006167.
GN Name=5730511K23Rik;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Oosada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
RA Hashimoto K.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB041615; BAA95098.1; -;
DR MGD; MGI:1923988; 5730511K23Rik.
DR InterPro; IPR007133; Pfam.
DR Pfam; PF03985; Pfam; 1.
SQ SEQUENCE 535 AA; 60534 MW; 6D7EBB1ECC8C075 CRC64;

Query Match 98.0%; Score 2708; DB 2; Length 535;
Best Local Similarity 97.9%; Pred. No. 2,1e-120;
Matches 524; Conservative 2; Mismatches 5; Indels 4; Gaps 1;
QY 1 MAPTTQQAQREDGHRPNSHRTLPSRGVGVCRVKYCNLSLPDIPDPKPTTYPFDQNRVQ 60
DB 1 MAPTTQQAQREDGHRPNSHRTLPSRGVGVCRVKYCNLSLPDIPDPKPTTYPFDQNRVQ 60
QY 61 YKATSLKQKHDLITLTPDLGVITIDLINPDTYRIDPNVLLPADKLEEEIQAPTSSKR 120
DB 61 YKATSLKQKHDLITLTPDLGVITIDLINPDTYRIDPNVLLPADKLEEEIQAPTSSKR 120
QY 121 SQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTFFTEEYIKRDSQITAE 180
DB 121 SQHAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTFFTEEYIKRDSQITAE 180
QY 181 KTFEDAQKSIHQYXSKPRVTPVEVMPVFPDFQWNPVCAQVIFDSDPAKPTSGAAALEM 240
DB 181 KTFEDAQKSIHQYXSKPRVTPVEVMPVFPDFQWNPVCAQVIFDSDPAKPTSGAAALEM 240
QY 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRKDOEEEMDYAPDDVDYKIDARENNVVK 300
DB 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRKDOEEEMDYAPDDVDYKIDARENNVVK 300
QY 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRKAQVQSGTNALLVVKHRDMNEKE 360
DB 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRKAQVQSGTNALLVVKHRDMNEKE 360
QY 361 LEAQARKAQLNHPPEEEEEEMEAEEKKAGGSEDEQEKSGSSEKSGSEDEHSGSESDR 420
DB 361 LEAQARKAQLNHPPEEEEEEMEAEEKKAGGSEDEQEKSGSSEKSGSEDEHSGSESDR 420
QY 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDSDDEHQAGGSDN 480
DB 421 BEGDRDEASDKSGGEDESSEDEARAARDKEEIFGSDADSDDDSDDEHQAGGSDN 480

```
QY 481 DSDSGNGGQOR-----SRSHRSRSPPPSGSHSAQEDGSEAAASDSSEADSDSD 531
Db 481 DSDSGSDGGQORSGSRSRSRSPPPSGSHSAQEDGSEAAASDSSEADSDSD 535

RESULT 5
O75239 PRELIMINARY; PRT; 510 AA.
AC O75239 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE F23149.1
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Gaines J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andrease T., Frankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Irong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005239; AAC25503.1; --
DR InterPro; IPR007133; Pfam1.
DR Pfam; PF03985; Pfam1.
SQ SEQUENCE 510 AA; 57456 MW; CACE73EDC7290CE8 CRC64;

Query Match 91.6%; Score 2531.5; DB 2; Length 510;
Best Local Similarity 91.7%; Pred. No. 4.4e-112;
Matches 498; Conservative 0; Mismatches 0; Indels 45; Gaps 3;

QY 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRFFVQ 60
Db 1 MAPIQTQAQREDG-----RSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRFFVQ 50

QY 61 YKATSLKQKHDLTPEDLGVTTIDLPDVTYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Db 51 YKATSLKQKHDLTPEDLGVTTIDLPDVTYRIDPNVLLDPADEKLEEEIQAPTSKR 110

QY 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITALE 180
Db 111 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITALE 170

QY 181 KTFEDAQKSIQSHYSKPRVTPVEVMPFPDPMWNPVPCAVIPDSDPAPKDTSGAAALEM 240
Db 171 KTFEDAQKSIQSHYSKPRVTPVEVMPFPDPMWNPVPCAVIPDSDPAPKDTSGAAALEM 207

QY 241 MSQAMIRGMMDEEENOFVAYFLPVEETLKVRKRDQEEEMDYPDDVDYKIRAREYNWVK 300
Db 208 MSQAMIRGMMDEEENOFVAYFLPVEETLKVRKRDQEEEMDYPDDVDYKIRAREYNWVK 267

QY 301 NKASKGYEENYFTIFREGDGVYVNELETR-----VRLSKRAKAGVQSGTNAL 348
Db 268 NKASKGYEENYFTIFREGDGVYVNELETRYSYSAHSYLSLDLVRLSKRAKAGVQSGTNAL 327

QY 349 LVYKHRDMKEKELEQAKAQLNHEPPEEEETETEKEAGGSDERQEGKSSEKEG 408
Db 328 LVYKHRDMKEKELEQAKAQLNHEPPEEEETETEKEAGGSDERQEGKSSEKEG 387

QY 409 SEDEHSGSSEREEDGRDASDCKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSD 468
Db 388 SEDEHSGSSEREEDGRDASDCKSGGEDESSEDEARAARDKEEIFGSDADSEDDADSD 447

QY 469 EDGQAGGSDNDSGSGNGGQSRSHRSRSPPPSGSHSAQEDGSEAAASDSSEADS 528
Db 448 EDGQAGGSDNDSGSGNGGQSRSHRSRSPPPSGSHSAQEDGSEAAASDSSEADS 507
```

```
QY 529 DSD 531
Db 508 DSD 510

RESULT 6
Q6P2Y1 PRELIMINARY; PRT; 520 AA.
AC Q6P2Y1;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein MGC76249.
GN Name=MGC76249;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Maman A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC064253; AAH64253.1; --
DR InterPro; IPR007133; Pfam1.
DR Pfam; PF03985; Pfam1.
KW Hypothetical protein.
SQ SEQUENCE 520 AA; 59064 MW; 76D526C423C459A7 CRC64;

Query Match 79.0%; Score 2183.5; DB 2; Length 520;
Best Local Similarity 80.2%; Pred. No. 1.2e-95;
Matches 429; Conservative 40; Mismatches 45; Indels 21; Gaps 8;

QY 1 MAPIQTQAQREDGHRPNSHRTLPERSGVVCRVKYCNLSLPDIPDPKFTYPPDQNRFFVQ 60
Db 1 MAPIQTQAQREDGHRSSSHRTLPERSGVVCRVKYCNLTLPDIPDPKFTYPPDQNRFFVQ 60

QY 61 YKATSLKQKHDLTPEDLGVTTIDLPDVTYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Db 61 YKATSLKQKHDLTPEDLGVTTIDLPDVTYRIDPNVLLDPADEKLEEEIQAPTSKR 120

QY 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITALE 180
Db 121 SQQAKVVPWVRKTEYISTEFNRYGISNEKPEVKIGSVKQQTTEEIYKDRDSQITALE 180

QY 181 KTFEDAQKSIQSHYSKPRVTPVEVMPFPDPMWNPVPCAVIPDSDPAPKDTSGAAALEM 240
Db 181 KTFEDAQKSIQSHYSKPRVTPVEVMPFPDPMWNPVPCAVIPDSDPAPKDTSGAAALEM 240
```



```

Qy 241 MSQAMIRGMDDEGNQFVAYFLPVETILKGRKRDQEEEMDYAPDDVDYDKIAREYNNVK 300
Db 241 MSQAMIRGMDDEGNQFVAYFLPGEITWRKRDQEEGLDYPEDYDKIAREYNNVK 300
Qy 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALLVVVHRDMNEKE 360
Db 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALLVVVHRDMNEKE 360
Qy 361 LEAQEARQAQLENHEPEEEEEEMETEKEEKGAGS 394
Db 361 LEAQEARQAQLENHEPEEEEEEMETEKEEKGAGS 394
Qy 351 LEAQEARQAQLENHEPEEEEEEMETEKEEKGAGS 384
Db 351 LEAQEARQAQLENHEPEEEEEEMETEKEEKGAGS 384

RESULT 8
Q9CS63 PRELIMINARY; PRT; 377 AA.
ID Q9CS63
AC Q9CS63;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mus musculus 8 days embryo whole body cDNA, RIKEN full-length enriched
DE library, clone:5730511K23 product:PD2 PROTEIN (HYPOTHETICAL 60.0 kDa
DE PROTEIN) homolog (Fragment).
GN Name=5730511K23Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishii K., Kiteunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,

```

```

Qy 241 MSQAMIRGMDDEGNQFVAYFLPVETILKGRKRDQEEEMDYAPDDVDYDKIAREYNNVK 300
Db 241 MSQAMIRGMDDEGNQFVAYFLPGEITWRKRDQEEGLDYPEDYDKIAREYNNVK 300
Qy 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALLVVVHRDMNEKE 360
Db 301 NKASKGYEENYFFIFREGDGVYVNELETRVRLSKRRKAGVQSGTNALLVVVHRDMNEKE 360
Qy 361 LEAQEARQAQLENHEPEEEEEEMETEKEEKGAGS 394
Db 361 LEAQEARQAQLENHEPEEEEEEMETEKEEKGAGS 394
Qy 351 LEAQEARQAQLENHEPEEEEEEMETEKEEKGAGS 384
Db 351 LEAQEARQAQLENHEPEEEEEEMETEKEEKGAGS 384

RESULT 7
Q8N7H5 PRELIMINARY; PRT; 485 AA.
ID Q8N7H5
AC Q8N7H5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein FLJ25557.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Suzuki Y., Hata H.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK098423; BAC05305.1; -.
DR InterPro: IPR007133; Pfam.
DR Pfam; PF03985; Pf1; 1
SQ SEQUENCE 485 AA; 55501 MW; 5F4A1ACC99142CID CRC64;

Query Match 72.2%; Score 1995; DB 2; Length 485;
Best Local Similarity 97.2%; Pred. No. 9.3e-87;
Matches 383; Conservative 1; Mismatches 0; Indels 10; Gaps 1;

Qy 1 MAPTIQQAQREDGHRPNSHHTLPSRGVCRVKYCNLSLPIDPDPKFIYFPQNRFPVQ 60
Db 1 MAPTIQQAQREDG-----RSGVCRVKYCNLSLPIDPDPKFIYFPQNRFPVQ 50

Qy 61 YKATSLKQKHQHDILTEPDGLVGTIDLPNTYRIDPNVLLDPADEKLEEEIQAPTSKR 120
Db 51 YKATSLKQKHQHDILTEPDGLVGTIDLPNTYRIDPNVLLDPADEKLEEEIQAPTSKR 110

Qy 121 SQQAKVVPWPKTEYISTENRYGINSKEPEVKIGSVKQKQFTEETIKDRDSQITAE 180
Db 111 SQQAKVVPWPKTEYISTENRYGINSKEPEVKIGSVKQKQFTEETIKDRDSQITAE 170

Qy 181 KTFEDAQKISQHSKSRVTPVPMVPPDPKMNINCAQVIFPDSDPAPKDTSGAAALEM 240
Db 171 KTFEDAQKISQHSKSRVTPVPMVPPDPKMNINCAQVIFPDSDPAPKDTSGAAALEM 230

Qy 241 MSQAMIRGMDDEGNQFVAYFLPVETILKGRKRDQEEEMDYAPDDVDYDKIAREYNNVK 300
Db 231 MSQAMIRGMDDEGNQFVAYFLPVETILKGRKRDQEEEMDYAPDDVDYDKIAREYNNVK 290

```

```
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Kohno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tegami M., Teraoka Y., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muranatsu M., Hayashizaki Y.;
RL Submittted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK017762; BAB30513.1; -
DR MGD: MGI:1923988; 5730511K23rik.
DR InterPro: IPR007133; Pfam.
DR Pfam: PF03985; Pfam; 1.
KW Hypothetical protein.
FT NON TER 377
SQ SEQUENCE 377 AA; 43836 MW; 4CE00D24EF5CEA CRC64;

Query Match 71.8%; Score 1984; DB 2; Length 377;
Best Local Similarity 100.0%; Pred. No. 2.3e-86;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPIQTQQRDGHRRPNSHRTLPERSGVVCVKYCNLSLPDIPDPKFTTYPDQNRVQ 60
Db |||||
1 MAPIQTQQRDGHRRPNSHRTLPERSGVVCVKYCNLSLPDIPDPKFTTYPDQNRVQ 60

QY 61 YKATSLKQKHGDLITPEPLGVITDILNPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Db |||||
61 YKATSLKQKHGDLITPEPLGVITDILNPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120

QY 121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGSVKQOFTBEIYKDRDSQITALE 180
Db |||||
121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGSVKQOFTBEIYKDRDSQITALE 180

QY 121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGSVKQOFTBEIYKDRDSQITALE 180
Db |||||
121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGSVKQOFTBEIYKDRDSQITALE 180

QY 181 KTFEDAQKSIQHYSKRPVTPVEVMPVDFPKWNPINCAQVIFDSDPAKPTSGAAALEM 240
Db |||||
181 KTFEDAQKSIQHYSKRPVTPVEVMPVDFPKWNPINCAQVIFDSDPAKPTSGAAALEM 240

QY 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNNVK 300
Db |||||
241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNNVK 300

QY 301 NKASKGYEENYFFTFREGDGVYVNELETRVRLSKRKAKAGVQSGTNALLVVKHRDMNEKE 360
Db |||||
301 NKASKGYEENYFFTFREGDGVYVNELETRVRLSKRKAKAGVQSGTNALLVVKHRDMNEKE 360

QY 361 LEAQEARKAQLNHEPE 377
Db |||||
361 LEAQEARKAQLNHEPE 377

RESULT 9
Q68F51 PRELIMINARY; PRT; 407 AA.
AC Q68F51;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE LOC446278 protein (Fragment).
GN Names:LOC446278;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
```

```
RT initiative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
EX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska J., Small D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RL Klein S., Gerhard D.S.;
RL Submittted (AUG-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC079993; AAH79993.1; -
DR InterPro: IPR007133; Pfam.
DR Pfam: PF03985; Pfam; 1.
FT NON TER 407
SQ SEQUENCE 407 AA; 47154 MW; 6CE32A7307186F83 CRC64;

Query Match 70.0%; Score 1935; DB 2; Length 407;
Best Local Similarity 90.0%; Pred. No. 5.2e-84;
Matches 367; Conservative 22; Mismatches 13; Indels 4; Gaps 2;

QY 1 MAPIQTQQRDGHRRPNSHRTLPERSGVVCVKYCNLSLPDIPDPKFTTYPDQNRVQ 60
Db |||||
1 MAPIQTQQRDGHRRPNSHRTLPERSGVVCVKYCNLSLPDIPDPKFTTYPDQNRVQ 60

QY 61 YKATSLKQKHGDLITPEPLGVITDILNPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120
Db |||||
61 YKATSLKQKHGDLITPEPLGVITDILNPDTYRIDPNVLLDPADEKLEEEIQAPTSSKR 120

QY 121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGSVKQOFTBEIYKDRDSQITALE 180
Db |||||
121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGSVKQOFTBEIYKDRDSQITALE 180

QY 121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGSVKQOFTBEIYKDRDSQITALE 180
Db |||||
121 SQQAKVVPWVRKTEYISTEFNRYGSGNEKPEVKIGSVKQOFTBEIYKDRDSQITALE 180

QY 181 KTFEDAQKSIQHYSKRPVTPVEVMPVDFPKWNPINCAQVIFDSDPAKPTSGAAALEM 240
Db |||||
181 KTFEDAQKSIQHYSKRPVTPVEVMPVDFPKWNPINCAQVIFDSDPAKPTSGAAALEM 240

QY 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNNVK 300
Db |||||
241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNNVK 300

QY 241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNNVK 300
Db |||||
241 MSQAMIRGMDDEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDYKIAREYNNVK 300

QY 301 NKASKGYEENYFFTFREGDGVYVNELETRVRLSKRKAKAGVQSGTNALLVVKHRDMNEKE 360
Db |||||
301 NKASKGYEENYFFTFREGDGVYVNELETRVRLSKRKAKAGVQSGTNALLVVKHRDMNEKE 360

QY 361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGSD-EEQEKGSSSEXE 407
Db |||||
361 LEAQEARKAQLNHEPEEEEEEMETEEKEAGSD-EEQEKGSSSEXE 407

RESULT 10
Q68F51 PRELIMINARY; PRT; 538 AA.
ID Q68F51
```

AC Q9VNS5; 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE CG2503-PA (LD37523p).
 GN ORFNames=CG2503;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426066; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.B.,
 RA Abriel J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gough J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.P., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,
 RA Nelson K.A., Nelson K.A., Nixon K., Nuskern D.R., Paclele J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler P., Shen H.,
 RA Shue B.C., Siden-Kimosh I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong E., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.O.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Dime E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Paclele J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
 melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
 RA Patel S., Fried E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 a genomics perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).

RN SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Friese E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Lieb G.,
 RA Miranda A., Mungall C.J., Nunco J., Paclele J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.E.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003605; AAF52095.1; -;
 DR EMBL; AY070561; AAL48032.1; -;
 DR FlyBase; FBgn0037274; CG2503.
 DR InterPro; IPR007133; Paf1.
 DR Pfam; PF03985; Paf1; 1.
 SQ SEQUENCE 538 AA; 60794 MW; D55E95B4F4EB8E51 CRC64;
 Query Match 45.0%; Score 1244.5; DB 2; Length 538;
 Best Local Similarity 50.0%; Pred. No. 3e-51;
 Matches 271; Conservative 66; Mismatches 172; Indels 38; Gaps 11;
 Qy 1 MAPIQTQQAQREDGHRPNHSRTLPERSGVVCRVKYCNLSLDPIDPDPKFIITYPFDONRFVQ 60
 Db 1 MPPTINNSAVNSAABK-RPQOTERKSEIICRVKYGNNLPDIPDLKPLQPFDSHFRVQ 59
 Qy 61 YKATSLKQKHDLITBPDLGVTIDLINPDTYRIDPNVLLDPADKLEELIQAPTSQR 120
 Db 60 YNFTSLERNFKYDLVLTEDLGVTVDLINRELYQADSWTLDPADKLEELITPTDSVR 119
 Qy 121 SQQAKVVPWVRKTEYISTEFNRYGISN-EKPEVKIGSVKQQTBEIYKDRDSQITAI 179
 Db 120 SRQHSRTVSWLRKSEYISTEQTRFPQPNLENIAKVGYNVKKSLREETLYLDREAQIKAI 179
 Qy 180 EKTPEDAQSISQHSYKSPRTVPEVMPVFPDKMWINPCAQVIFDSDPAPKDTSGAAALE 239
 Db 180 EKTFSDTKSIITHYKSPNVVPEVLFDPFTNWKFPQVIFDSDPAPAKNVPAQLE 239
 Qy 240 MMSQAMIRGMWDEGNQFVAYFYPVEETLKKRKDQSEEMDYAPDDVYDYKIAREYNWV 299
 Db 240 EMSQAMIRGMWDESGEQFVAYFPTETQLEKRTDFINGELYKEEEVEYKIAREYNWV 299
 Qy 300 KNKASKGYEENYFFIFREGDGVYNNLETRVLSKRRAKAGVQSGTQNALLVVKHRDMMNK 359
 Db 300 KTKASKGYEENYFFVFMQ-DGIYNELETRVLRNKRVRKVG-QQPNNTKLVVYKRPDLSM 357
 Qy 360 ELEQAEARKAQLNHEPEER-----EEEM-----ETEE-----KENGSD----- 395
 Db 358 EHRMQRVRRQLEVPGESEIEVEEVEEQMIIIGETEKTSDDAAGQAASGADSPAQV 417
 Qy 396 --EQEKGSGSSEKSGSESEEREEDGRDSDKSGSESEDESEDESEDESEDESEDESEDE 453
 Db 418 ARDQSRSRRTSGS--SSGSGSGSGSRASRSKSGSGSGSGSRSTNSPAGSKSGSR- 475

QY 454 FGSADSEDDADDDDEBGOACGSDNDSDGS-NGGQSRSHSRSSASPPFSGSEHSAQ 512
Db 476 SRVSRSRSKSGSRSRSGSRSGSRSGSRSGSRSGSRSGSRSGSRSGSRSGSRSGSA 535
QY 513 ED 514
Db 536 SD 537

RESULT 11
Q7PXA3
ID Q7PXA3 PRELIMINARY; PRT; 468 AA.
AC Q7PXA3;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE AGCP12185 (Fragment).
GN Names=agCG49165; ORFNames=ENSANGG00000017836;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008987; EAA01198.1; -.
DR InterPro; IPR007133; Pfam1.
DR Pfam; PF03985; Pfam1; 1.
FT NON TER 468 468
FT NON TER 468 468
SQ SEQUENCE 468 AA; 53298 MW; 47BECB1BB520C826 CRC64;

Query Match 40.8%; Score 1129; DB 2; Length 468;
Best Local Similarity 61.8%; Pred. No. 7.4e-46;
Matches 223; Conservative 45; Mismatches 87; Indels 6; Gaps 5;

QY 1 MAPIQTQAOQREDGHRPNSHRTLPERSGVCRVKYCNLSLPDIPDPKFTITYPFDQNRVQ 60
Db 5 MAPIVQNGANGAD-KRP--VRQERSELISRVKYNLTLPDIPDLKFTITYPFDNRFIQ 61
QY 61 YKATSLKQKHDLTEPDLGVTIDINPTYRIDPNVLLDPADEKLEEEIOAPTSSKR 120
Db 62 YNPTSLENYRYEVLTEHDLGVTIDINRDLYQIDHSAQLDPADEKLEEDIHPTQDSMR 121
QY 121 SQQHAKVVPWPKRTEYISTEFNRYGISN-EKPEVKIGVSVKQOFTTEELYKDRDSOITAI 179
Db 122 SSRHAKSVSLRSEYISTEQTFNPTQWEKVEAKYGFNVKSLREETLYMDREAQIKAI 181
QY 180 EKTFFDAQKSIQYKSKRPVTPVEVMPVFPDFKWINPACQVIFDSDPAKDTSGAAALE 239
Db 182 EKTFFEDTKPITTHYSKPGVTPVEINPVPDFANWKYPCAQVIFDSDPAPSKNVAQIE 241
QY 240 MMSQAMIRGMMDSEGNQFVAYFLPVEETLKRRKRDQEEEMDYAPDDVYDKIAREYNWV 299
Db 242 EMSQAMIRGMVDSGEQFVAYFLPTDITLKRRDLVNETLYDEDEEYEQWAREYNWV 301
QY 300 KMKASKGYEENYFFI PREGDGVYVNELETFRVLSKRAKAGVQSGTNALLVVKHRDMNEK 359
Db 302 KSKASKGYEENYVTLR-P-DGIYVNELETFRVLSKRRQK-NAQQQSNITKLWVKHRPLNAS 359
QY 360 E 360
Db 360 E 360

RESULT 12
P90783

ID P90783 PRELIMINARY; PRT; 425 AA.
AC P90783;
DT 01-MAY-1997 (T-EMBLrel. 03, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein C55A6.9.
GN ORFNames=C55A6.9;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Kershaw J.K.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z81051; CAB02869.1; -.
DR PIR; T20261; T20261.
DR WormBase; WBGene00008338; C55A6.9.
DR WormPep; C55A6.9; CR20614.
DR InterPro; IPR007133; Pfam1.
DR Pfam; PF03985; Pfam1; 1.
KW Hypothetical protein.
SQ SEQUENCE 425 AA; 49897 MW; 4BA06AF90913DCC5 CRC64;

Query Match 23.3%; Score 645; DB 2; Length 425;
Best Local Similarity 33.1%; Pred. No. 5e-23;
Matches 146; Conservative 96; Mismatches 165; Indels 34; Gaps 10;

QY 24 PERSGVCRVKYCNLSLPDIPDPKFTITYPFDQNRVQYKATSLKQKHHLTEPDLG 82
Db 14 PRKVDPMKPRFTNTVPDPAKFWTCFVPLGRFVFQPAAYRDYKGVICDDMDGL 73
QY 83 TIDLNPITYRIDP-NVLLDPADEKLEEEIOAPTSSKRSQAHAKVVPWPKRTEYISTEF 141
Db 74 NVDLIDLKKYDEDPTEIDEKONILLDDGAAKLIARSSQHSKLVPMWPKRTEYISTEF 133
QY 142 NRYGISNKPVKIGVSVKQOFTTEELYKDRDSOITAEKTFEDAKTSISCHYKPRVTP 201
Db 134 NRGVTDARQRTKGLYNLKKQQVEDMYRDKSQSDAINKTFEDVPRKPVKHYKGVKA 193
QY 202 VEVMPVFPDFKWINPACQVIFDSDPAKDTSGAAALEMMSQAMIRGMMDSEGNQFVAYF 261
Db 194 VEESVFPDFDHNKHLFAHVQFDGDTITTEFEEDERQOARESSVIKAMEREDQKFAAVF 253
QY 262 LPVEETLKRRKRDQEEEMDYAPDDVYDKIAREYNWVKNKASKGYEENYFFI PREGDGV 321
Db 254 VPTIGCLTSFMDLELERPFDEDMKYEFLLSREYTFKMEHLPPR--DRDVFIMYHRNNVF 311
QY 322 YVNELETFRVLSKRAKAGVQSGTNALLVVKHRDMNEKELEAQAQKALQHNHE-PEEE 380
Db 312 QYNEVDCNVKMT-RKPRMALSRSKSLTYR-----NPSELEQKQNMKREAHLYEQKTRK 366
QY 381 EEEME--TEEKAGSGDEEQKSGSSEKSGSEGESEHSGSEEREEDRDEASDKSGSGE 438
Db 367 QEILEKIQEKEEGD-----SSDQSSDDDDKPKSR-----SDSSSDV 406
QY 439 SSEDARAARDKEEIFGSDAD 459
Db 407 SSDDD--SPRKKEPTVSDSD 425

RESULT 13
Q8RW91
ID Q8RW91 PRELIMINARY; PRT; 589 AA.

Q8RW91.
AC 01-JUN-2002 (TREMELrel. 21, Created)
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
DE At1g79730/F19K16.29.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kaniya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan Y.W.,
RA Ishida J., Jones T., Kaniya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY093780; AM103971; -;
DR EMBL; BT002711; AA011627.1; -;
DR InterPro; IPR007133; Pfam1.
DR Pfam; PF03985; Pfam1.
SQ SEQUENCE 589 AA; 67336 MW; 49D9F81P59C43AFF CRC64;

Query Match 13.5%; Score 372.5; DB 2; Length 589;
Best Local Similarity 25.5%; Pred. No. 5.4e-10;
Matches 127; Conservative 78; Mismatches 152; Indels 141; Gaps 20;

QY 6 QTQQRD-GHR---PNSHRT-----LPERSGVVCRVKYCN 37
DB 143 ELEKQKQDEKHKRQMKNSHKSQMPKGTBEKPTPLTTDRVQNRKLPKTFICKLFRN 202
QY 38 SLDPDPKPFITVPQNRVVOYKATSLKQKHDLTPDLGVTIDLNPTRIDPN 97
DB 203 ELDPSPQALKMTIKKDKQDFTKYITSLKLPKPFVFPDLGIPDLDDLVSYN-PPK 261
QY 98 VL--LDPADKLEEEIQAPTSRSQOQAKVVP-----WMRKTEYISTEFNRYGISNE 149
DB 262 VKAPLAPADEELLRDD--DAVTPKDGIRRKERTDKGMSLWLVKTQYISS--INNE 314
QY 150 KPEVIGVSVKQOFTBEI-----YKDRDSQITAEIKTFPDAOKSISQHY 194
DB 315 -----SARQSLTEKQAKELREMGGINILHNNRERQIKDIEASF-ACKSRPVHA 365
QY 195 SKPRVTVEVMPVPDFPKWNPINCAQVIFDS-----DPAPKDTSGAALWMS 242
DB 366 TNKNLQVEVLPPLPYFDRDEQVVFANFDGAPIADSEFFGKLDPSIRDAHESRAL--LK 423
QY 243 QAMIRGMDBEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDKIAREYNWVKNK 302
DB 424 SYVVGSDTANPEKFLAYMVPSPDLSDKIDHENEISYT-----WVREYLDVQPN 475
QY 303 ASKGYENYFFIPREGDGVYNELETVRLSKRAKAGVQSG-----TNA 347
DB 476 AND--PCTLVSPDNSTASYL-PLPMLNLKRRAREGRSSDEIHEFPVPSRVTVRRST 532
QY 348 LVVVKGRD--MNEKLEAQARKAQLEN-----HEPEEBEEMETEKEAGSD 396
DB 533 VSVIEHKDSGVYSRVGASSKMRRLDEGLGRSWKHEPEQD----- 575

QY 397 EQKGSSESKEGSEDBHS 414
DB 576 -----ANQYSDGNEDDYS 588

RESULT 14
Q9MA04 PRELIMINARY; PRT; 593 AA.
AC Q9MA04;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE F20B17.16.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chio J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei Q., Chin C., Chio J., Choi E.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010793; AAF68118.1; -;
DR InterPro; IPR007133; Pfam1.
DR Pfam; PF03985; Pfam1.
SQ SEQUENCE 593 AA; 68056 MW; 3E188CF6A2P8E61A CRC64;

Query Match 13.1%; Score 361.5; DB 2; Length 593;
Best Local Similarity 25.8%; Pred. No. 1.8e-09;
Matches 130; Conservative 73; Mismatches 152; Indels 149; Gaps 21;

QY 6 QTQQRD-GHR---PNSHRT-----LPERSGVVCRVKYCN 37
DB 143 ELEKQKQDEKHKRQMKNSHKSQMPKGTBEKPTPLTTDRVQNRKLPKTFICKLFRN 202
QY 38 SLDPDPKPFITVPQNRVVOYKATSLKQKHDLTPDLGVTIDLNPTRIDPN 91
DB 203 ELDPSPQALKMTIKKDKQDFTKYITSLKLPKPFVFPDLGIPDLDDLVS 262
QY 92 YRIDPNVL--LDPADKLEEEIQAPTSRSQOQAKVVP-----WMRKTEYISTEFN 143
DB 263 YN-PPKVPKAPLAPADEELLRDD--DAVTPKDGIRRKERTDKGMSLWLVKTQYISS-- 316
QY 144 YGISNEKPVKIGSVKQOFTBEI-----YKDRDSQITAEIKTFPDAOK 188
DB 317 ---INNE-----SARQSLTEKQAKELREMGGINILHNNRERQIKDIEASF-ACK 365
QY 189 STSQSVSKPRVTVEVMPVPDFPKWNPINCAQVIFDS-----DPAPKDTSGAA 236
DB 366 SPVVRATNKQVLPPLPYFDRDEQVVFANFDGAPIADSEFFGKLDPSIRDA 421
QY 237 ALEMMSQAMIRGMDBEGNQFVAYFLPVEETLKKRKRDQEEEMDYAPDDVYDKIAREYN 296
DB 422 HESRVSVELPISMNTANPEKFLAYMVPSPDLSDKIDHENEISYT-----WVREYL 473
QY 297 WNVKNKASKGYENYFFIPREGDGVYNELETVRLSKRAKAGVQSG----- 344

```
Db 474 WDVQPNAND--PGTYLVSPDNGTASYL-PLPMRLNLKRRAREGRSSDEIEHFPVPSRVT 530
Qy 345 ----TNALLVVKHRD--MNEKLELAQEAQKAQLEN-----HPEPEEEEEEMETEKE 390
Db 531 VRRRSTVSVIEHKDGSYSSRVGASSSKWRLLEDEGLGRSNKHEPEQD----- 579
Qy 391 AGSDDEEQEGSSSEKESSEDEHS 414
Db 580 -----ANQYSDGNEDDYS 592

RESULT 15
Q6ZD92 PRELIMINARY: PRT; 451 AA.
AC Q6ZD92;
DT 05-JUL-2004 (TremBLrel 27, Created)
DT 05-JUL-2004 (TremBLrel 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel 27, Last annotation update)
DE Proline-rich protein-like.
GN Name=P043808.25;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004460; BAC99509.1; -.
DR InterPro; IPR007133; Paf1.
DR Pfam; PF03985; Paf1; 1.
SQ SEQUENCE 451 AA; 52159 MW; 297158E4B56642AE CRC64;

Query Match 13.0%; Score 358.5; DB 2; Length 451;
Best Local Similarity 26.9%; Pred. No. 1.8e-09;
Matches 115; Conservative 80; Mismatches 180; Indels 53; Gaps 17;

Qy 17 PMSHR---TLPSRGVVCVKYCNSLPDIPDPKFTITYPFDQNRVQYKATSLKQHKHD 73
Db 47 PNAERFENLKKPTPLCKHKFRELDPSPSSQLKWLPLNKKDKRYTKYRITSLEKNYIPK 106
Qy 74 LITEPDLGVTIILNPDTYRIDP-NVLLDPADEKLL-EEEIQAPTS-----KRSQOHAKV 127
Db 107 MIVPEDLGIPDLLQMSVNTFPVQPMAPDEDEELLRDDDEVLTVPKKGIRKERTDKG 166
Qy 128 VPWMKTEYI---STEFNRYGI----SNEKPEVKIGSVKQQTBEIEYKDRDSQITAE 180
Db 167 MSWLVTQYIISLSTDAKMSITEKQAKERRSREG--RNTFLEN--INDREKQIKALE 221
Qy 181 KTFEDAQKISQHSYKRPVTPVEVMVPFPDFKXWINPCAQVIFDSDPAPKDTSGAAALE- 239
Db 222 DSFR-AAKSRPVHOTKRGMEAEWVLPPLPDFORYDDQFVMVNFDDGPT-ADSEQYNKLER 279
Qy 240 -----MMSQAMIRGMDEEGNQFVAYFLPVEETLKRRKQDEEMDYAPDDVVDY 289
Db 280 SERDECESAVMKGSFLVNGSDPAKQEKFLAYMVPSPHLSKOLDDETDIQYS----- 332
Qy 290 KIAREYNMNVKNAKSGYEENFFIFREGDGVVYNELETRVRLSKRAKAGVQSGTNALL 349
Db 333 -WLREYHWEVRGD-DKDDPTTYLVTF-DDDGAKYLPPTKLVLOKKKAKEG-RSGDE--- 385
Qy 350 VVKHRDMNKELELAQEAQKAQLENHPEPEEEEEEMETEKEAGGSDDEQEGSSSEKES 409
Db 386 -IEHFPVPSRITVSRTHAGGMMEHGESSNMH-----ENLKRQRSSVDDDLIDHPKHSRVED 440
Qy 410 EDEHSGSE 417
Db 441 MDQYSGDE 448
```

Search completed: September 26, 2005, 13:54:48
Job time : 183 secs